(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 June 2003 (05.06.2003)

PCT

(10) International Publication Number
WO 03/046126 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/34001

(22) International Filing Date: 23 October 2002 (23.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/335,115	24 October 2001 (24.10.2001)	US
10/140,545	7 May 2002 (07.05.2002)	US
60/391,367	25 June 2002 (25.06.2002)	US

(71) Applicant: **AMERSHAM BIOSCIENCES (SV) CORP** [US/US]; 928 East Arques Avenue, Sunnyvale, CA 94085 (US).

(72) Inventors: **SAMARTZIDOU, Hrissi**; 938G La Mesa Terrace, Sunnyvale, CA 94086 (US). **HOUTS, Thomas**; 8485 Burchell Road, Gilroy, CA 95020 (US). **YANG, Wen**; 2025 California Street, #30, Mountain View, CA 94040 (US). **BUI, Son**; 1243 Stellar Way, Milpitas, CA 95035 (US). **HARKINS, Timothy**; 884 Myrtle Street, San Jose, CA 95126 (US).

(74) Agent: **RONNING, Royal, N., Jr.**; Amersham Biosciences Corp, 800 Centennial Avenue, Piscataway, NJ 08855 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/046126 A2

(54) Title: ARTIFICIAL GENES FOR USE AS CONTROLS IN GENE EXPRESSION ANALYSIS SYSTEMS

(57) Abstract: Method of producing universal controls for use in gene expression analysis systems such as macroarrays, real-time PCR, northern blots, SAGE and microarrays. The controls are generated either from near-random sequence of DNA, or from intergenic or intronic regions of a genome. Twenty-three specific control sequences are also disclosed. Also presented are methods of using these controls, including as negative controls, positive controls, and as calibrators of a gene expression analysis system.

ARTIFICIAL GENES FOR USE AS CONTROLS IN GENE EXPRESSION ANALYSIS SYSTEMS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of United States patent application number 10/140,545, filed May 7, 2002, which claims priority to United States provisional patent application number 60/289,202, filed May 7, 2001, and 60/312,420, filed August 15, 2001. This application also claims priority to United States provisional patent application serial number 60/335,115, filed October 24, 2001, and 60/391,367, filed June 25, 2002, the disclosures of which are incorporated herein by reference in their entireties.

REFERENCE TO SEQUENCE LISTING SUBMITTED ON COMPACT DISC

The present application includes a Sequence Listing filed on one CD-R disc, provided in duplicate, containing a single file named pto_PB0181.txt, having 56 kilobytes, last modified on October 21, 2002, and recorded on October 21, 2002. The Sequence Listing contained in said file on said disc is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

1. Field of the Invention:

The present invention relates to a method of using artificial genes as universal controls in gene expression analysis systems. More particularly, the present invention relates to a method of producing universal Controls for use

in gene expression analysis systems such as macroarrays, real-time PCR, northern blots, SAGE and microarrays, such as those provided in the Microarray ScoreCard system.

2. Description of Related Art:

Gene expression profiling is an important biological approach used to better understand the molecular mechanisms that govern cellular function and growth. Microarray analysis is one of the tools that can be applied to measure the relative expression levels of individual genes under different conditions. Microarray measurements often appear to be systematically biased, however, and the factors that contribute to this bias are many and ill-defined (Bowtell, D.L., *Nature Genetics* 21, 25-32 (1999); Brown, P.P. and Botstein, D., *Nature Genetics* 21, 33-37 (1999)). Others have recommended the use of "spikes" of purified mRNA at known concentrations as controls in microarray experiments. Affymetrix includes several for use with their GeneChip products. In the current state of the art, these selected genes are actual genes selected from very distantly related organisms. For example, the human chip (designed for use with human mRNA) includes control genes from bacterial and plant sources.

Each of the prior art controls consists of transcribed sequences of DNA from some source. As a result, that source cannot be the subject of a hybridization experiment using those controls due to the inherent hybridization of the controls to its source. In addition, the lack of universal references consistent from experiment to experiment and from species to species greatly reduces the ability for scientists to compare data across labs, users, or time. What is needed, therefore, is a set of universal controls that do not hybridize with the DNA of any

source which may be the subject of an experiment. More desirably, there is a need for a universal control for gene expression analysis which do not hybridize with any known source.

SUMMARY OF THE INVENTION

Accordingly, this invention provides a process of producing universal controls that are useful in gene expression analysis systems designed for any species and which can be tested to insure lack of hybridization with mRNA from sources other than the control DNA itself.

The invention relates in a first embodiment to a process for producing at least one universal control for use in a gene expression analysis system. The process comprises selecting at least one non-transcribed (preferably intergenic, also intronic) region of genomic DNA from a known sequence, designing primer pairs for said at least one non-transcribed region and amplifying said at least one non-transcribed region of genomic DNA to generate corresponding double stranded DNA, then cloning said double stranded DNA using a vector to obtain additional double stranded DNA and formulating at least one control comprising said double stranded DNA.

The present invention relates in a second embodiment to a process of producing at least one universal control for use in a gene expression analysis system wherein testing of said at least one non-transcribed region to ensure lack of hybridization with mRNA from sources other than said at least one non-transcribed region of genomic DNA is performed.

The present invention in a third embodiment relates to said process further comprising purifying said DNA and mRNA, determining the concentrations thereof and

formulating at least one control comprising said DNA or of said mRNA at selected concentrations and ratios.

Another embodiment of the present invention is a universal control for use in a gene expression analysis system comprising a known amount of at least one DNA generated from at least one non-transcribed region of genomic DNA from a known sequence, or comprising a known amount of at least one mRNA generated from DNA generated from at least one non-transcribed region of genomic DNA from a known sequence. The present invention may optionally include generating mRNA complementary to said DNA and formulating at least one control comprising said mRNA, by optionally purifying said DNA and mRNA, determining the concentrations thereof and formulating at least one control comprising said DNA or of said mRNA at selected concentrations and ratios.

Another embodiment of the present invention is a universal control for use in a gene expression analysis system wherein a known amount of at least one DNA sequence generated from at least one non-transcribed region of genomic DNA from a known sequence, a known amount of at least one mRNA generated from DNA generated from at least one non-transcribed region of genomic DNA from a known sequence is included, and the aforementioned control wherein, said DNA and mRNA do not hybridize with any DNA or mRNA from a source other than the at least one non-transcribed region of genomic DNA.

The present invention, relates to a method of using said universal control, as a negative control in a gene expression analysis system by adding a known amount of said control containing a known amount of DNA, to a gene expression analysis system as a control sample and subjecting the sample to hybridization conditions in the

absence of complementary labeled mRNA and examining the control sample for the absence or presence of signal.

Further, said controls can be used in a gene expression analysis system by adding a known amount of a said control containing a known amount of DNA to a gene expression analysis system as a control sample and subjecting the sample to hybridization conditions, in the presence of a said control containing a known amount of labeled complementary mRNA, and measuring the signal values for the labeled mRNA and determining the expression level of the gene transcript based on the signal value of the labeled mRNA.

Additionally, said controls may be used as calibrators in a gene expression analysis system by adding a known amount of a said control containing known amounts of several DNA sequences to a gene expression analysis system as control samples and subjecting the samples to hybridization conditions in the presence of a said control containing known amounts of corresponding complementary labeled mRNAs, each mRNA being at a different concentration and measuring the signal values for the labeled mRNAs and constructing a dose-response or calibration curve based on the relationship between signal value and concentration of each mRNA.

Also, the present invention relates to a method of using said controls as calibrators for gene expression ratios in a two-color gene expression analysis system by adding a known amount of at least one of said controls containing a known amount of DNA to a two-color gene expression analysis system as control samples and subjecting the samples to hybridization conditions in the presence of a said control containing known amounts of two differently labeled corresponding complementary labeled mRNAs for each DNA sample present and measuring the ratio of the signal

values for the two differently labeled mRNAs and comparing the signal ratio to the ratio of concentrations of the two or more differently labeled mRNAs.

A further embodiment of the present invention is a process of producing controls that are useful in gene expression analysis systems designed for any species and which can be tested to insure lack of hybridization with mRNA from sources other than the synthetic sequences of DNA from which the control is produced.

One or more such controls can be produced by a process comprising synthesizing a near-random sequence of non-transcribed DNA, designing primer pairs for said at least one near random sequence and amplifying said non-transcribed DNA to generate corresponding double stranded DNA, then cloning said double stranded DNA using a vector to obtain additional double stranded DNA and formulating at least one control comprising said double stranded DNA.

The process can also be used to produce at least one control for use in a gene expression analysis system wherein testing of said sequence of non-transcribed synthetic DNA to ensure lack of hybridization with mRNA from sources other than said sequence of non-transcribed DNA is performed.

Additionally, mRNA complementary to said synthetic DNA can be generated and formulated to generate at least one control comprising said mRNA.

DNA and mRNA can be subsequently purified, the concentrations thereof determined, and one or more controls comprising said DNA or said mRNA at selected concentrations and ratios be formulated.

Another embodiment of the present invention is a control for use in a gene expression analysis system produced by the process comprises synthesizing a near-random sequence of DNA, designing primer pairs for said synthetic

DNA and amplifying said DNA to generate corresponding double stranded DNA, then cloning said double stranded DNA using a vector to obtain additional double stranded DNA and formulating at least one control comprising a known amount of at least one said double stranded DNA or a known amount of at least one mRNA generated from said DNA, and optionally, wherein, said DNA and mRNA do not hybridize with any DNA or mRNA from a source other than said DNA sequence of non-transcribed DNA.

The present invention, additionally, relates to a method of using said controls containing a known amount of DNA, as a negative control in a gene expression analysis system including adding a known amount of said control containing a known amount of DNA to a gene expression analysis system as a control sample, and subjecting the sample to hybridization conditions in the absence of complementary labeled mRNA and examining the control sample for the absence or presence of signal.

Further, said controls may be used in a gene expression analysis system wherein a known amount of a said control containing a known amount of DNA is added to a gene expression analysis system as a control sample and subjecting the sample to hybridization conditions in the presence of a said control containing a known amount of labeled complementary mRNA and measuring the signal values for the labeled mRNA and determining the expression level of the gene transcript based on the signal value of the labeled mRNA.

The present invention, also relates to a method of using said controls as calibrators in a gene expression analysis system including adding known amounts of a said control containing known amounts of several DNAs to a gene expression analysis system as control samples and subjecting the samples to hybridization conditions in the presence of a

said control containing known amounts of corresponding complementary labeled mRNAs, each mRNA being at a different concentration and measuring the signal values for the labeled mRNAs and constructing a dose-response or calibration curve based on the relationship between signal value and concentration of each mRNA.

The present invention, additionally, relates to a method of using said controls as calibrators for gene expression ratios in a two-color gene expression analysis system comprising adding a known amount of at least one of said controls containing a known amount of DNA to a two-color gene expression analysis system as control samples and subjecting the samples to hybridization conditions in the presence of a said control containing known amounts of two differently labeled corresponding complementary labeled mRNAs for each DNA sample present and measuring the ratio of the signal values for the two differently labeled mRNAs and comparing the signal ratio to the ratio of concentrations of the two or more differently labeled mRNAs.

Further embodiments and uses of the current invention will become apparent from a consideration of the ensuing description.

BRIEF DESCRIPTION OF THE DRAWINGS

The above and other objects and advantages of the present invention will be apparent upon consideration of the following detailed description taken in conjunction with the accompanying drawings, in which like characters refer to like parts throughout, and in which:

FIG. 1 shows representative results for the selection of universal controls that do not cross-hybridize with human RNA;

FIG. 2 shows representative results for the selection of universal controls that do not cross-hybridization with each other;

FIG. 3 represents a performance evaluation of the universal controls;

FIG. 4 shows a scatter plot of raw signals for the calibration and ratio controls from a two-color hybridization experiment;

FIG. 5 shows calibration curves based on the Calibration controls for a representative hybridization experiment;

FIG. 6 presents the control nucleotide sequence of DR1 (SEQ ID NO: 1);

FIG. 7 presents the control nucleotide sequence of DR2 (SEQ ID NO: 2);

FIG. 8 presents the control nucleotide sequence of DR3 (SEQ ID NO: 3);

FIG. 9 presents the control nucleotide sequence of DR4 (SEQ ID NO: 4);

FIG. 10 presents the control nucleotide sequence of DR5 (SEQ ID NO: 5);

FIG. 11 presents the control nucleotide sequence of DR6 (SEQ ID NO: 6);

FIG. 12 presents the control nucleotide sequence of DR7 (SEQ ID NO: 7);

FIG. 13 presents the control nucleotide sequence of DR8 (SEQ ID NO: 8);

FIG. 14 presents the control nucleotide sequence of DR9 (SEQ ID NO: 9);

FIG. 15 presents the control nucleotide sequence of DR10 (SEQ ID NO: 10);

FIG. 16 presents the control nucleotide sequence of RC1 (SEQ ID NO: 11);

FIG. 17 presents the control nucleotide sequence of RC2 (SEQ ID NO: 12);

FIG. 18 presents the control nucleotide sequence of RC3 (SEQ ID NO: 13);

FIG. 19 presents the control nucleotide sequence of RC4 (SEQ ID NO: 14);

FIG. 20 presents the control nucleotide sequence of RC5 (SEQ ID NO: 15);

FIG. 21 presents the control nucleotide sequence of RC6 (SEQ ID NO: 16);

FIG. 22 presents the control nucleotide sequence of RC7 (SEQ ID NO: 17);

FIG. 23 presents the control nucleotide sequence of RC8 (SEQ ID NO: 18);

FIG. 24 presents the control nucleotide sequence of Utility1 (SEQ ID NO: 19);

FIG. 25 presents the control nucleotide sequence of Utility2 (SEQ ID NO: 20);

FIG. 26 presents the control nucleotide sequence of Utility3 (SEQ ID NO: 21);

FIG. 27 presents the control nucleotide sequence of Negative1 (SEQ ID NO: 22);

FIG. 28 presents the control nucleotide sequence of Negative2 (SEQ ID NO: 23);

FIG. 29 presents the nucleotide sequence of DR1s used in a spike mix (SEQ ID NO: 24);

FIG. 30 presents the nucleotide sequence of DR2s used in a spike mix (SEQ ID NO: 25);

FIG. 31 presents the nucleotide sequence of DR3s used in a spike mix (SEQ ID NO: 26);

FIG. 32 presents the nucleotide sequence of DR4s used in a spike mix (SEQ ID NO: 27);

FIG. 33 presents the nucleotide sequence of DR5s used in a spike mix (SEQ ID NO: 28);

FIG. 34 presents the nucleotide sequence of DR6s used in a spike mix (SEQ ID NO: 29);

FIG. 35 presents the nucleotide sequence of DR7s used in a spike mix (SEQ ID NO: 30);

FIG. 36 presents the nucleotide sequence of DR8s used in a spike mix (SEQ ID NO: 31);

FIG. 37 presents the nucleotide sequence of DR9s used in a spike mix (SEQ ID NO: 32);

FIG. 38 presents the nucleotide sequence of DR10s used in a spike mix (SEQ ID NO: 33);

FIG. 39 presents the nucleotide sequence of RC1s used in a spike mix (SEQ ID NO: 34);

FIG. 40 presents the nucleotide sequence of RC2s used in a spike mix (SEQ ID NO: 35);

FIG. 41 presents the nucleotide sequence of RC3s used in a spike mix (SEQ ID NO: 36);

FIG. 42 presents the nucleotide sequence of RC4s used in a spike mix (SEQ ID NO: 37);

FIG. 43 presents the nucleotide sequence of RC5s used in a spike mix (SEQ ID NO: 38);

FIG. 44 presents the nucleotide sequence of RC6s used in a spike mix (SEQ ID NO: 39);

FIG. 45 presents the nucleotide sequence of RC7s used in a spike mix (SEQ ID NO: 40);

FIG. 46 presents the nucleotide sequence of RC8s used in a spike mix (SEQ ID NO: 41);

FIG. 47 presents the nucleotide sequence of Utility1s used in a spike mix (SEQ ID NO: 42);

FIG. 48 presents the nucleotide sequence of Utility2s used in a spike mix (SEQ ID NO: 43);

FIG. 49 presents the nucleotide sequence of Utility3s used in a spike mix (SEQ ID NO: 44);

FIG. 50 presents the nucleotide sequence of Negative1s used in a spike mix (SEQ ID NO: 45); and

FIG. 51 presents the nucleotide sequence of Negative2s used in a spike mix (SEQ ID NO: 46).

DETAILED DESCRIPTION OF THE INVENTION

The present invention teaches universal Controls for use in gene expression analysis systems such as microarrays. Many have expressed interest in being able to obtain suitable genes and spikes as controls for inclusion in their arrays.

An advantage of the universal Controls of this invention is that a single set can be used with assay systems designed for any species, as these Controls will not be present unless intentionally added. This contrasts with the concept of using genes from "distantly related species." For example, an analysis system directed at detecting human gene expression might employ a *Bacillus subtilis* gene as control, which may not be present in a human genetic material. But this control might be present in bacterial genetic material (or at least, cross hybridize), thus it may not be a good control for an experiment on bacterial gene expression. The novel universal Controls presented here provide an advantage over the state of the art in that the same set of controls can be used without regard to the species for the test sample RNA.

The present invention employs the novel approaches of using either non-transcribed genomic sequences or totally random synthetic sequences as a template and generating both DNA and complementary "mRNA" from such sequences, for use as controls. The Controls could be devised *de novo* by designing near-random sequences and synthesizing them resulting in synthetic macromolecules as universal controls. Totally synthetic random DNA fragments are so designed that they do not cross-hybridize with each other or with RNA from

any biologically relevant species (meaning species whose DNA or RNA might be present in the gene expression analysis system). The cost of generating such large synthetic DNA molecules can be high. However, they only need to be generated a single time. Additionally, fragment size can be increased by ligating smaller synthetic fragments together by known methods. In this way, fragments large enough to be easily cloned can be created. Through cloning and PCR sufficient quantities of DNA for use as controls can be produced and mRNA can be generated by *in vitro* transcription for use in controls.

A simpler approach is to identify sequences from the intergenic or intronic regions (referred here as non-transcribed regions) of genomic DNA from an organism, and use these as a template for synthesis via PCR (polymerase chain reaction). Ideally, sequences of around 1000 bases (could range from 500 to 2000 bases) are selected based on computer searches of publicly accessible sequence data. The criteria for selection include:

1. The sequence must be from a non-transcribed region; and
2. The sequence must not have homology with or be predicted to hybridize with any known / published gene or expressed sequence tag (EST).

PCR primer pairs are designed for the selected sequence(s) and PCR is performed using genomic DNA (as a template) to generate PCR fragments (double strand DNA) corresponding to the non-transcribed sequence(s) as the control DNA. Additional control DNA can be cloned using a vector and standard techniques. Subsequently, standard techniques such as *in vitro* transcription are used to generate mRNA (complementary to the cDNA and containing a poly-A tail) as the control mRNA. Standard techniques are

used for purifying the Control DNA and Control mRNA products, and for estimating their concentrations.

Empirical testing is also performed to ensure lack of hybridization between the Control DNA on the array and other mRNAs, as well as with mRNA from important gene expression systems (e.g., human, mouse, *Arabidopsis*, etc.).

The above approaches were used to generate twenty-three universal control sequences from intergenic regions of the yeast *Saccharomyces cerevisiae* genome. Specifically, using yeast genome sequence data publicly available (http://genome-www.stanford.edu/Saccharomyces/), intergenic regions approximately 1 kb in size were identified. These sequences were BLAST'd and those showing no homology to other sequences were identified as candidates for artificial gene controls. Candidates were analyzed for GC-content and a subset with a GC-content of ≥36% was identified. Specific primer sequences have been identified and primers synthesized. PCR products amplified with the specific primers have been cloned directly into the pGEM™-T Easy vector (Promega Corp., Madison, WI). Both array targets and templates for spike mRNA have been amplified from these clones using distinct and specific primers.

A greater number of intergenic regions have been cloned for testing. DNA samples from all the candidates were amplified, spotted on glass microarray slides and hybridized with mRNA samples from several species and each candidate spike mRNA, respectively, to identify those that do not cross-hybridize. First, they were screened for no cross-hybridization with RNA from different biological species. mRNA from human (eight tissues: skeletal muscle, spleen, liver, heart, kidney, brain, placenta and lung), mouse (six tissues: skeletal muscle, spleen, liver, heart, kidney and brain), rat (six tissues: skeletal muscle, spleen, liver, heart, kidney and brain), yeast (*S.*

cerevisiae) and bacteria (*E. coli* and two Archaea species), as well as total RNA from plant (*Arabidopsis*, Oil Palm) were tested against the control candidates. Candidates that did not cross-react with the RNA samples from the species tested were then selected for cross-hybridization with each other. The candidates were hybridized with each candidate mRNA independently.

From the candidate clones that exhibited specific hybridization, twenty-three were included into the final set of universal controls. FIG. 6 through FIG. 28 presents the nucleotide sequences of the twenty-three controls spotted on the microarray slides, while FIG. 29 through 51 presents the nucleotide sequences of the twenty-three controls that were transcribed and used in a spike mix, respectively. SEQ ID NO: 1 through SEQ ID NO: 23 present the nucleotide sequences of the twenty-three controls spotted on the microarray slides, while SEQ ID NO: 24 through SEQ ID NO: 46 present the nucleotide sequences of the controls that were transcribed and used in a spike mix.

These universal controls, when included in microarray experiments, perform as:

1. Negative controls: Control DNA included in the array, but for which no complementary artificial mRNA is spiked into the RNA sample, serves as a negative control;
2. Calibration controls: Several different Control DNA samples may be included in an array, and the complementary Control mRNA for each is included at a known concentration, each having a different concentration of mRNA. The signals from the array features corresponding to these Controls or Calibrators may be used to construct a "dose-response curve" or calibration curve to

estimate the relationship between signal and amount of mRNA from the sample;

3. Ratio controls: In two-color microarray gene expression studies, it is possible to include different, known, levels of Control mRNA complementary to Control DNA in the labeling reaction for each channel. The ratio of signals for the two dyes from a particular gene can be compared to the ratio of signals from the two dyes of the Control mRNA. This can serve as a test of the accuracy of the system for determining gene expression ratios.
4. Utility controls: These controls can be added into the sample preparation steps (such as RNA extraction and purification) for normalization of the biological samples and assessment of sample losses during preparation. Alternatively, they can be added to labeling reactions as additional calibrators or ratios.

Mixtures of several different Control mRNA species can be prepared (spike mixes) at known concentrations and ratios to simplify and standardize the experimental protocol while providing a comprehensive set of precision and accuracy information. Table 1 demonstrates one embodiment of this concept. The mRNA from the final set of clones have been pre-mixed at specific concentrations and ratios so they can serve as the various controls when hybridized to their corresponding control DNA spotted on the arrays. Ten calibrators (those included in the labeling reaction at a ratio of 1:1) spanning a dynamic range of 4.5 orders of magnitude are included as calibration controls. Eight ratio controls are included, at two expression levels (low and medium to high) and reversed with respect to the reference and test samples.

The universal controls as shown in Table 1 can be used as references for microarray validation and standardization across biological species and experimental platforms. These controls can be used to verify the accuracy and precision of gene expression ratios, and the sensitivity and dynamic range of the microarray system. Through the use of Calibration (standard) curves, these controls may allow reporting gene expression levels in consistent mass units, improving the comparisons of results across laboratories.

The following examples demonstrate how these Control DNA and Control mRNA were generated, and then used as universal controls in microarray gene expression experiments. They are representative of the many different types of experiments that could benefit from the use of these controls. The following examples are offered by way of illustration and not by way of limitation.

Table 1. Suggested Control mRNA spike mix composition for two-color gene expression ratio experiments.

Control Type	Control Name	Target Cy3:Cy5 Ratio	mRNA in the Spike Mix (pg/2µl of spike)	
			Cy3	Cy5
Calibration	DR1s	1:1	30 000	30 000
Calibration	DR2s	1:1	10 000	10 000
Calibration	DR3s	1:1	3 000	3 000
Calibration	DR4s	1:1	1 000	1 000
Calibration	DR5s	1:1	300	300
Calibration	DR6s	1:1	100	100
Calibration	DR7s	1:1	30	30
Calibration	DR8s	1:1	10	10
Calibration	DR9s	1:1	3	3
Calibration	DR10s	1:1	1	1
Ratio	RC1s	3:1 low	300	100
Ratio	RC2s	1:3 low	100	300
Ratio	RC3s	3:1 high	3 000	1 000
Ratio	RC4s	1:3 high	1 000	3 000

Ratio	RC5s	10:1 low	300	30
Ratio	RC6s	1:10 low	30	300
Ratio	RC7s	10:1 high	10 000	1 000
Ratio	RC8s	1:10 high	1 000	10 000
Utility	utility1s	User defined	User defined	User defined
Utility	Utility2s	User defined	User defined	User defined
Utility	Utility3s	User defined	User defined	User defined
Negative	Negative1s	NA	0	0
Negative	Negative2s	NA	0	0

Example 1. Generation of Artificial Controls from Intergenic Regions of *S. cerevisiae* Genome.

Using yeast genomic sequence data publicly available (http://genome-www.stanford.edu/Saccharomyces/), intergenic regions (YIRs) approximately 1 kb in size were identified. These sequences were BLAST'd and those showing no homology to other sequences were identified as candidates for artificial gene controls. Candidates were analyzed for GC-content and a subset with a GC-content of ≥36% was identified. Specific primer sequences have been identified and synthesized. PCR products amplified with the specific primers have been cloned directly into the pGEM™-T Easy vector (Promega Corp., Madison, WI). Both array targets and templates for spike mRNA have been amplified from these clones using distinct and specific primers.

When used as DNA controls, the YIR sequences were amplified by PCR with specific primers, using 5 ng of cloned

template (plasmid DNA) and a primer concentration of 0.5μM in a 100 μl reaction volume, and cycled as follows: 35 cycles of 94°C 20 sec., 52°C 20 sec., 72°C 2 min., followed by extension at 72°C for 5 min.

All YIR control mRNAs for the spike mix are generated by *in vitro* transcription. Templates for *in vitro* transcription (IVT) are generated by amplification with specific primers that are designed to introduce a T7 RNA polymerase promoter on the 5' end and a polyT (T21) tail on the 3' end of the PCR products. Run-off mRNA is produced using 1 μl of these PCR products per reaction with the AmpliScribe system (Epicentre, Madison, WI). IVT products are purified using the RNAEasy system (Qiagen Inc., Valencia, CA) and quantified by spectrophotometry.

Initially, fifty intergenic region sequences have been cloned for testing. DNA samples from all the candidates were amplified, spotted on glass microarray slides and hybridized with mRNA samples from several species and each candidate spike mRNA, respectively, to identify those that do not cross-hybridize. First, they were screened for no cross-hybridization with RNA from different biological species. mRNA from human (8 tissues: skeletal muscle, spleen, liver, heart, kidney, brain, placenta and lung), mouse (6 tissues: skeletal muscle, spleen, liver, heart, kidney and brain), rat (6 tissues: skeletal muscle, spleen, liver, heart, kidney and brain), yeast (*S. cerevisiae)* and bacteria (*E. coli* and two Archaea species), as well as total RNA from plant (*Arabidopsis*, Oil Palm) were tested against the control candidates.

Figure 1 shows the hybridization of candidates with human brain mRNA. The results indicated that two YIR clones, 33 and 62, hybridized with human brain RNA while the other candidates did not (since no appreciable signal is detected). Clones, such as 33 and 62, that exhibited such

template (plasmid DNA) and a primer concentration of 0.5μM in a 100 μl reaction volume, and cycled as follows: 35 cycles of 94°C 20 sec., 52°C 20 sec., 72°C 2 min., followed by extension at 72°C for 5 min.

All YIR control mRNAs for the spike mix are generated by *in vitro* transcription. Templates for *in vitro* transcription (IVT) are generated by amplification with specific primers that are designed to introduce a T7 RNA polymerase promoter on the 5′ end and a polyT (T21) tail on the 3′ end of the PCR products. Run-off mRNA is produced using 1 μl of these PCR products per reaction with the AmpliScribe system (Epicentre, Madison, WI). IVT products are purified using the RNAEasy system (Qiagen Inc., Valencia, CA) and quantified by spectrophotometry.

Initially, fifty intergenic region sequences have been cloned for testing. DNA samples from all the candidates were amplified, spotted on glass microarray slides and hybridized with mRNA samples from several species and each candidate spike mRNA, respectively, to identify those that do not cross-hybridize. First, they were screened for no cross-hybridization with RNA from different biological species. mRNA from human (8 tissues: skeletal muscle, spleen, liver, heart, kidney, brain, placenta and lung), mouse (6 tissues: skeletal muscle, spleen, liver, heart, kidney and brain), rat (6 tissues: skeletal muscle, spleen, liver, heart, kidney and brain), yeast (*S. cerevisiae)* and bacteria (*E. coli* and two Archaea species), as well as total RNA from plant (*Arabidopsis*, Oil Palm) were tested against the control candidates.

Figure 1 shows the hybridization of candidates with human brain mRNA. The results indicated that two YIR clones, 33 and 62, hybridized with human brain RNA while the other candidates did not (since no appreciable signal is detected). Clones, such as 33 and 62, that exhibited such

cross-hybridization were removed from the set of candidates for universal controls.

Candidates that did not cross-react with the RNA samples from the species tested were then tested for cross-hybridization with each other. The candidates were hybridized with each candidate mRNA independently. In Figure 2 the labeled mRNA made from clone #50 was specifically hybridized against all other candidate clones. It hybridized only to its corresponding target DNA and can be included into the candidate set. However, clone #52 bound to the spot of clone #49 besides its own and therefore was not included in the candidate set.

From the candidate clones that exhibited specific hybridization, twenty-three are included into the final set of universal controls. FIG. 6 through FIG. 28 presents the nucleotide sequences of the twenty-three controls spotted on the microarray slides, while FIG. 29 through 51 presents the nucleotide sequences of the twenty-three controls as used in a spike mix, respectively. The sequences of these clones are further presented in the Sequence Listing, incorporated herein by reference in its entirety, as follows:

SEQ ID NOs: 1 - 23 (nt, control nucleotide sequences, including calibration controls 1 through 10, ratio controls 1 through 8, utility controls 1 through 3, and negative controls 1 and 2 respectively);

SEQ ID NOs: 24 - 46 (nt, spike mix nucleotide sequences, including calibration controls 1 through 10, ratio controls 1 through 8, utility controls 1 through 3, and negative controls 1 and 2 respectively);

Upon confirmation of the exact structure, each of the above-described nucleic acids of confirmed structure is recognized to be immediately useful as a control.

Example 2. Performance Evaluation of the Artificial Controls.

The universal controls (both the spike mixes and their corresponding spotting samples) have been evaluated for their performance in real microarray experiments and tested for the following.

Experimental design, including array design and the hybridization sample concentration were tested (Figure 3). Control samples were spotted in five replicates and hybridized with probes prepared with the spike mix only or the spike mixes with skeletal muscle mRNA. The same array image in Figures 3 is shown at two different gray scales for easy visualization of signals across the entire dynamic range.

Universal utility, including hybridization of the spikes on pre-arrayed slides from various species were also tested. The controls showed no cross-hybridization on human, rat, mouse, *Arabidopsis*, Yeast and *E. coli* pre-arrayed slides from commercial sources (data not shown).

Spike mix performance was tested, including ratio performance and Calibration curves (Figures 4 and 5). The mRNA from the final set of clones have been pre-mixed at specific concentrations and ratios (see Table 1 above) so they can serve as the various controls when hybridized to their corresponding control DNA spotted on the arrays. Ten calibrators (those included in the labeling reaction at a ratio of 1:1), spanning a dynamic range of 4.5 orders of magnitude, are included as calibration controls. Eight ratio controls are included, at two expression levels (low and

medium to high) and reversed with respect to the reference and test samples.

Figure 4 shows a scatter plot of raw signals for the calibration and ratio controls from a two-color hybridization experiment. The Calibrators are accurately and precisely clustered at the 45-degree line and the ratios at their expected target values at high (labeled 'H') and low (labeled 'L') levels of expression.

Figure 5 shows calibration curves based on the Calibration controls for a hybridization experiment. In this "standard curve", the Cy3 and Cy5 signals from the calibration controls are plotted as a function of the amount of mRNA in the spike mix. The error bars represent the 95% confidence intervals for the mean value. From such curves, attributes such as the limit of detection, the linear dynamic range and the signal saturation limit can be assessed. The application of the universal controls for the generation of standard curves can be the first step towards true quantitation of expression levels from microarray experiments.

The controls as shown in Table 1 can be used as references for microarray validation and standardization across biological species and experimental platforms. These controls can be used to verify the accuracy and precision of gene expression ratios, and the sensitivity and dynamic range of the microarray system. Through the use of Calibration (standard) curves, these controls may allow reporting gene expression levels in consistent mass units, improving the comparisons of results across laboratories

The above examples illustrate specific aspects of the present invention and are not intended to limit the scope thereof in any respect and should not be so construed.

Those skilled in the art having the benefit of the teachings of the present invention as set forth above, can

effect numerous modifications thereto. These modifications are to be construed as being encompassed within the scope of the present invention as set forth in the appended claims.

5

What is claimed is:

1. A control for use in a gene expression analysis system comprising:

(a) a known amount of at least one DNA selected from the group consisting of

(i) SEQ ID Nos: 1 - 23;

(ii) a degenerate variant of the sequence set forth in (i); and

(iii) a complement of the sequence set forth in (i) and (ii); or

(b) a known amount of at least one mRNA transcribed from the group consisting of

(i) SEQ ID NOs: 24 - 46;

(ii) a degenerate variant of the sequence set forth in (i); and

(iii) a complement of the sequence set forth in (i) and (ii).

2. A method of using a control as a negative control in a gene expression analysis system comprising:

adding a known amount of said control DNA of claim 1, to a gene expression analysis system as a control sample;

subjecting the sample to hybridization conditions in the absence of complementary labeled mRNA;

examining the control sample for the absence or presence of signal.

3. A method of using controls in a gene expression analysis system comprising:

adding a known amount of said control DNA of claim 1, to a gene expression analysis system as a control sample;

subjecting the sample to hybridization conditions in the presence of a known amount of labeled complementary mRNA of claim 1;

measuring the signal values for the labeled mRNA and determining the expression level of the DNA based on the measured signal value.

4. A method of using controls as calibrators in a gene expression analysis system comprising:

adding a known amount of a said control containing known amounts of several DNAs of claim 1, to a gene expression analysis system as control samples;

subjecting the samples to hybridization conditions in the presence of a said control containing known amounts of corresponding complementary labeled mRNAs of claim 1, each mRNA being at a different concentration;

measuring the signal values for the labeled mRNAs and constructing a dose-response or calibration curve based on the relationship between signal value and concentration of each mRNA.

5. A method of using controls as calibrators for gene expression ratios in a two-color gene expression analysis system comprising:

adding a known amount of at least one of said controls containing a known amount of DNA of claim 1, to a two-color gene expression analysis system as control samples;

subjecting the samples to hybridization conditions in the presence of a said control containing known amounts of two differently labeled corresponding complementary labeled mRNAs of claim 1, for each DNA sample present;

measuring the ratio of the signal values for the two differently labeled mRNAs and comparing the signal ratio to the ratio of concentrations of the two or more differently labeled mRNAs.

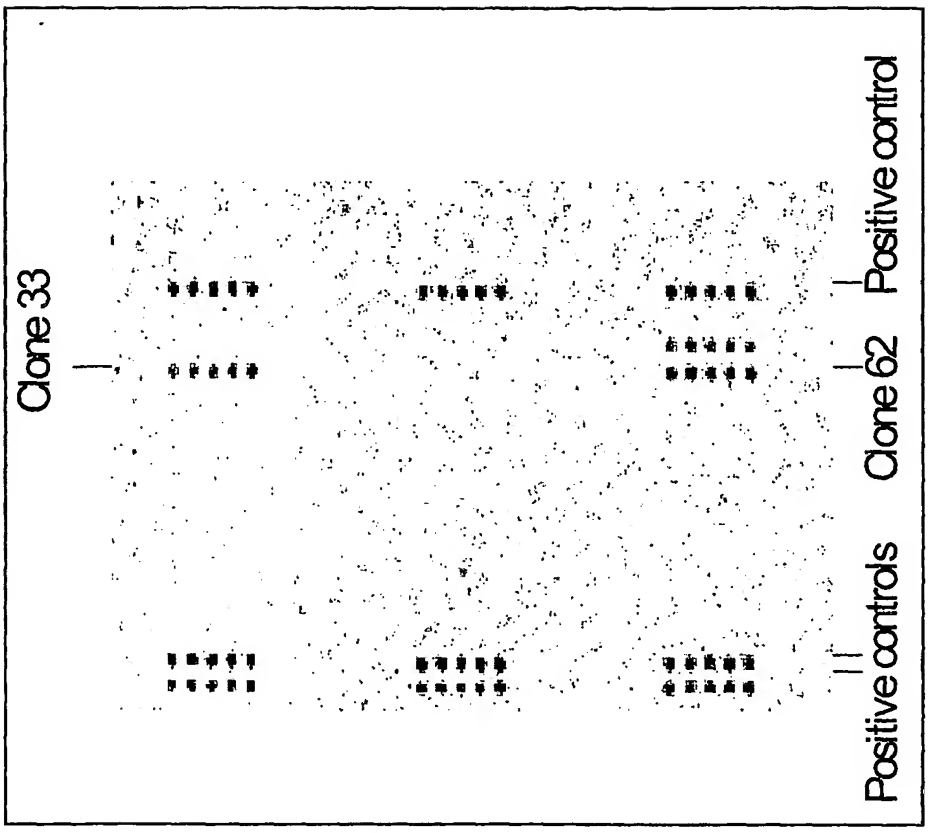

FIG. 1

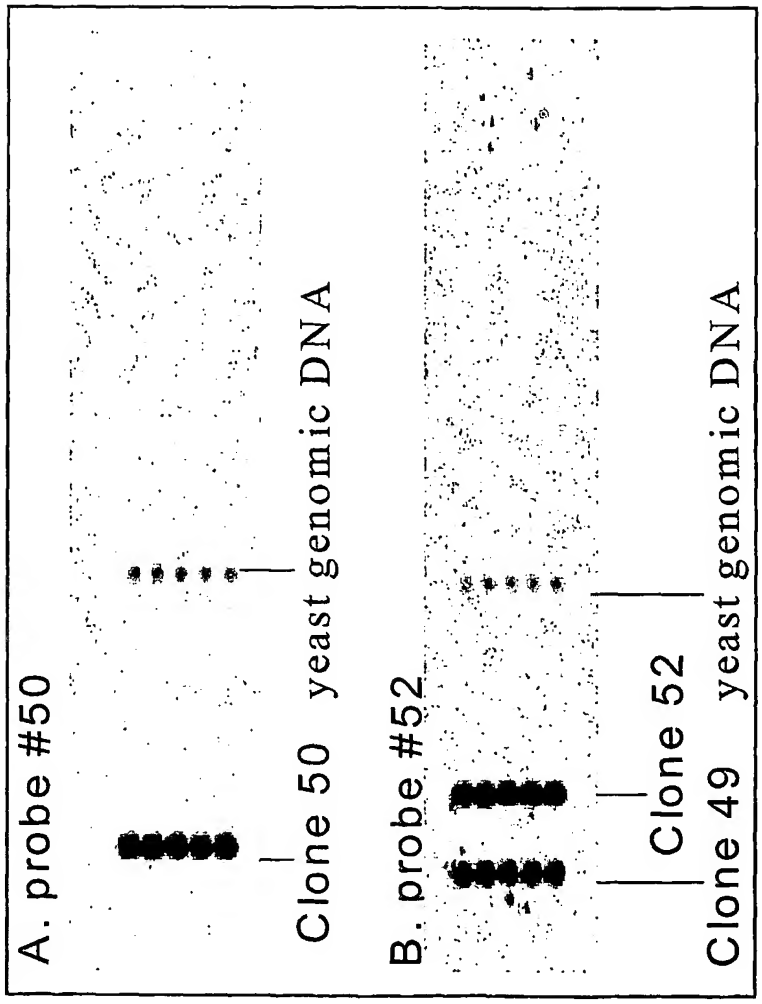
A. probe #50
Clone 50 yeast genomic DNA
B. probe #52
Clone 52
Clone 49 yeast genomic DNA

FIG. 2

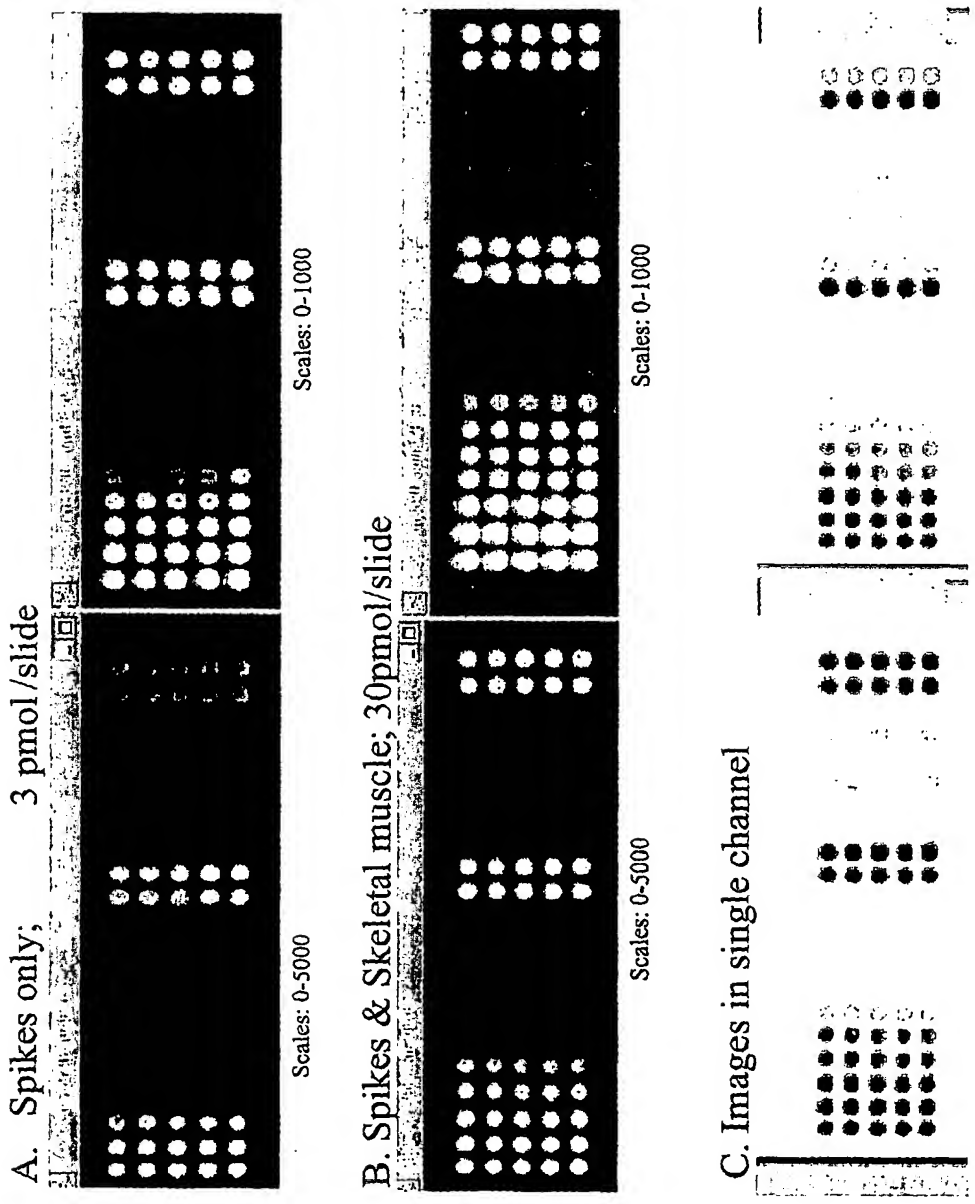
A. Spikes only; 3 pmol /slide
Scales: 0-5000
Scales: 0-1000
B. Spikes & Skeletal muscle; 30pmol/slide
Scales: 0-5000
Scales: 0-1000
C. Images in single channel
Cy5
Cy3

FIG. 3

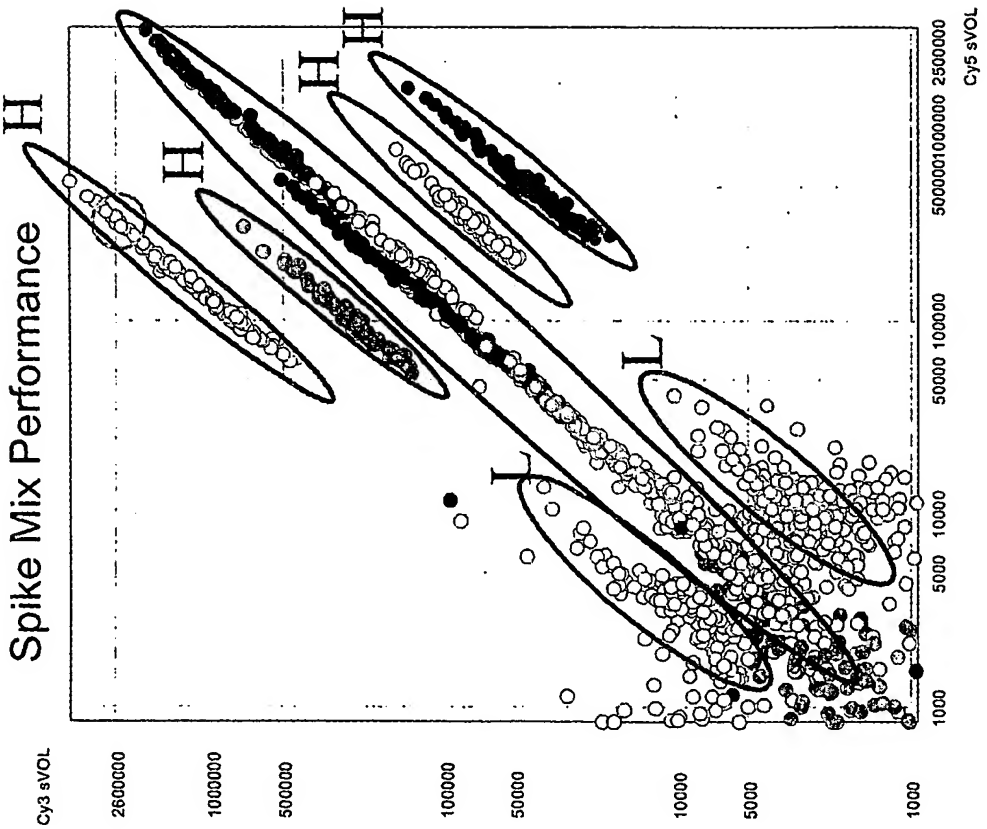
Spike Mix Performance
H
H
H
H
L
L
Cy3 sVOL
2600000
1000000
500000
100000
50000
10000
5000
1000
1000
5000
10000
50000
100000
500000
1000000
2500000
Cy5 sVOL

FIG. 4

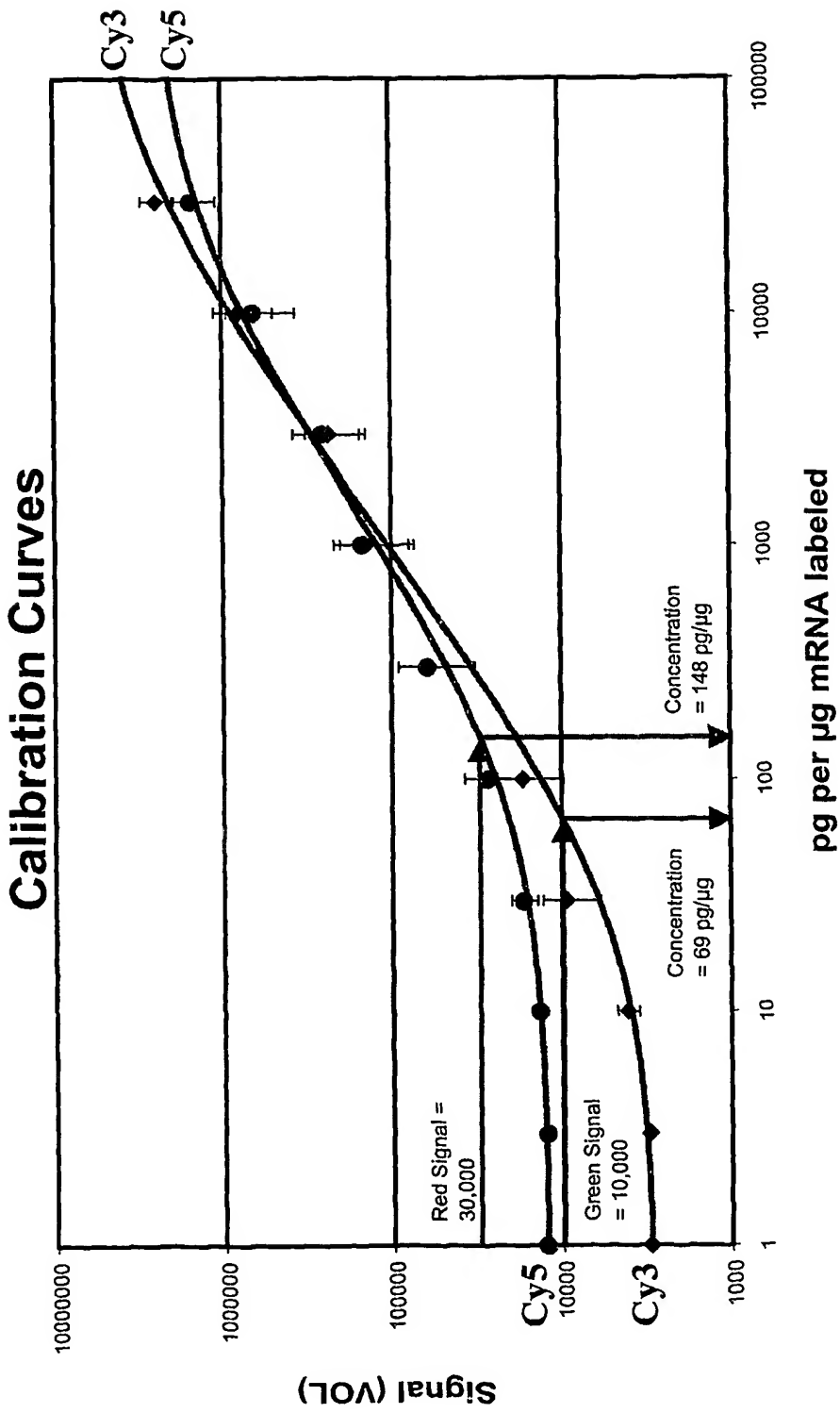
Calibration Curves
Cy3
Cy5
Signal (VOL)
pg per µg mRNA labeled
10000000
1000000
100000
10000
1000
1
10
100
1000
10000
100000
Red Signal = 30,000
Green Signal = 10,000
Concentration = 148 pg/µg
Concentration = 69 pg/µg
Cy5
Cy3

FIG. 5

DR1

nt: SEQ ID NO: 1

```
tgttgtccaagaaaggagggattttgttcatcagaaaagaattcagaaaa     50
gcaaggaaacagtactatcgtttagaatgtagaatgataggttgcttgct    100
aattctattatggcacgaatgatacacccatattttcaacaaaatcaata    150
cccactagcatcattgagccaactatttgtcaatgcaaccattaccggta    200
cttcatcctgatttaacgagtctacttttttatcacgtcaaaatttactt    250
gttttcctgtaaacccgaaataaaggcaaaaaagacctgggtgcaattac    300
gaataaatgtacaataatcatcctgtttgcatagtaaacttccagttaga    350
gtcacacaacgcaatgaattttgacagttttctgtgcgatattctttggt    400
aaacgtaaagaacaggcaacttttggtacaatggattctagcccatatgg    450
ttcatttctggtgcattcgcaaagtcagtatttgtctagctgtgttttct    500
ggctgagagacattatgatgttattcattgttatggatatctctgtagct    550
catgctgcttatttctccctaaaaaagtttttttctctcgaatacattctt    600
gaccatttcatagtgaaattcttgtacttatttaaaaccaaaaatggaag    650
tattcatacatccccctatcaaaaacactcaataagtttcgaattattcg    700
ttcgtctaaacagtgtccaatactcaaaggggtattcaagacggcacaaa    750
atcagcatcttcccttatccgtgttccagaaataccacgctaaggttttt    800
cctcctacaatccataaaatcattaaggaggcagcttgaaaaatcttgaa    850
attcaaaagagttatcttgggctaatcgaaattaacgataaccagagtag    900
aatattcaagatcacagctccaccttagtttcgagg                  936
```

FIG. 6

DR2

nt: SEQ ID NO: 2

```
gcgaataaccaaaacgagactacttttaccattacaaccattttctttt      50
tccctatttctcactggttgacagaaatcagtgtgctatcatcctaccat     100
atgcgctaaacttattgtctttctcctcctagagatgctgtattccatgc     150
atattctgaacgatgggttggtgtttttatcaagcaaggttaatcacatg     200
gcgtggcttgctccacacatcagtagaaaacgcataccgcagcggaatcc     250
ttaaataataagtgattttactgttcatcaactacaatcggactctttca     300
caattacccttcttgttttccacatttactgttaaatgaagggatgtaca     350
gaaggcttaggaaaacctgtgctgaatactggatggacactgcattccca     400
cagtgaaacttttatagatacactgtcagttattttcgaactttcatcaa     450
gttgctgagttttagtatccctttgccttagctatatgtttgaatgagca     500
aaatatttgcaatgtctctagctttcttgaaatattggtttatattgagg     550
gcttggtaagatttcaaatttcactttgaaatactcaggagaaaaatcat     600
gctcttttgataatttggtgactaaacatacataaaacagtttaattttg     650
ggtggtaatggctgtgtgactagctatagaaagaaaaaaattaaaaaaaa     700
aaaaaaaaatcaagtagttcctgcactgcgacgtccattatagcattatg     750
aattggtccctgatttacgcatgcgataaactatttttagcgcagccgca     800
tattatccgagaataacttccgacataagaaaattcgcagaaaatagata     850
aaaaactgctcttggcattcttcacttcctctattacacactgtgtcata     900
ccacaatcatctcacagtatgtatttgtatgtttatacatgctataacgt     950
aaaacaatgtagaatatatatctaaatacctcacggttttagtttagtgc    1000
```

FIG. 7

DR3

nt: SEQ ID NO: 3

```
tgtgagaaattcactagcttcacctaaaagcaaaaaatctcaaaattccc      50
aatattcacatagtctaaagtaccgatagcaaccaacatatataaacagt     100
agtattttacgaagctgaattgcaagattagtgagaggagaataaccgga     150
taatttttttggattacgttattgttaaaggctataatattaggtgaaac     200
agaatgtcctagaagtttttttctttcatgttaaatttattgattcttgc     250
gcttcagcttttataaaacataagaactgtttcttcacgttaacttcttg     300
tgccacatataatgatgtactagtaatatgggtactatttggcagatgat     350
atttgatttttattcaagacggttactgtttctacgattgatattttcat     400
tcctggatatcatcttgccagatcacttacaatttaggccgcgcctgaat     450
tgaagagtacttcaatacgtagtgtactgtccaaactctcttccaaattt     500
ttaatatttagctggggttgggtaacaagtgagcaagggaaaaagtgaac     550
attttaagaagaacaataaaatagcaagagatggaatggtaatgcttggc     600
tctcgagaagagtagcataaaacgagacttgtttaaaacaggatatgaca     650
tacttcaattcagctttccctatcagccgctcgagcagttatataggtgt     700
gttgccggagtaatttggcggaggccaacagtggctaggcggcaacgcct     750
ggaacacgcgcttaaaagttctggaaggttcgcgaattgagaactgctca     800
ggggcgaatacaggggcggccttggcggcagggggggaggcctctgtgaag     850
ttagttatataagacttgctgtcatcgtttttttgatcccggcaggaact     900
atcttttattctcatacatacggtcaagaagtataattatacataacata     950
gggacacgttcaggcaattgtccatatccc                          980
```

FIG. 8

DR4

nt: SEQ ID NO: 4

```
gatgtctgttttctttatgcaggatattaaatacaagtttgtgcttaaga      50
aatttccatgaagatatcaacatttattgtaggaattgcataaatagatg     100
aattatgtgccgctggacgtttatagatagcataagcacaatgacttaaa     150
ggttataatactcattgatatcactctgattataaaatcgtaatatgcga     200
ataggtgaactaatcggaataacccatacgacacttcaagcttcaattct     250
atttcaactgtagtgcctgctagtgaagaatacaaaagtagcatacgtga     300
tgtgcaaaaaatgcgctacttatcacacaagtaccttgcgcaagaagggt     350
actctaaaccggggccatcgcattaccagacggagatgtattctttatga     400
agcaataattggaggtgtatcaagttcgaaactgctgatgctatggattt     450
acatctttcttatgcacaaggcttgcttgtgtttctgagtagttagtttt     500
tagatttttgtcaagtctggggtaagttaattcgagcaaaattaacggca     550
cgttattctaatgcatatgttgttcatatattcttttacaaagaggtttg     600
gaatgatgtcaccgatgttagaatgttaggagaatttcatgtgaatttta     650
gtccaagtgttgaagttctcttctgcagttagggcacgtacatggcaacg     700
atatcgtttttgatgtattaatcttagtaggcgttgagtttgtatgttac     750
ttttctcaggtgatgaagcgtgatgacgatgacaaaaatgggttataata     800
gggcgcactatcatcatgcgtgattgatatttaaccaatgtcttgagtac     850
atcaactccagaaaatgggtcattatatgcctagcatgtattatttgaga     900
cataaagttttatctcgagaccttgacgtataggaaa                  937
```

FIG. 9

DR5

nt: SEQ ID NO: 5

```
ctgtagatttgcatggacgcacgtcgcccatacgccaaactttggcaatg      50
atactcgttattcgtaatatcagtccgtcaaggtgctgtgatttctctat     100
tttatattgcctattatttttcaaatgatttgagccgttttaaattgag      150
tatgcaatgagtcttttgaatcaaccgtaaggcagttccataaccactgc     200
cacgaatacgtttcactaccttgaagaatctctaatgtaggccgtattct     250
tcgcacttagttctgacgatgtagacatctcattatataagagcataagc     300
gcctgtttctagaatcatttcttcgtgacccagcttttgagttatttcg      350
cggtattttgaaacatttctcgagcttgacgtgaacatccttatatttca     400
tgacaaactcgatcattggaacatccctgcctcgattttagagctagtat     450
caaatttcaatctctttgtgatggagccccgctcctatttcaaaagagaa     500
gtttcttgtatgcatatgttattgaagtctgattatagcaagtgcaatgt     550
cgtctcaattattttaactatttttagccatacatgttagttatcctcaa     600
agagagcctccagactgggaagcagtgtttgtcatttcaaataagtagat     650
ttcacagtttgtatgattttcgaagccaggattcattgggctttgagtaa     700
agagaagccgcgtattacgaacagcttacgatattgtaaaatattccctt     750
attgtggtgccccaatggatacatgccagagaaatgtctgtgaaattgaa     800
caattacaatgacgagagcaagtaatccggcggccttgtctctctttcac     850
tagtaccgtctatatctcttgagcgccaatatgcgaaagctttcacaagg     900
ttgatgttcatggtattcggcgtcgatagcgaattgctta               940
```

FIG. 10

DR6

nt: SEQ ID NO: 6

ttagtttggaacagcagtgtagataccgtccttggatagagcgctggaga 50
tagctggtctcaatctggtggagtaccatgggacaccagtgatgactcta 100
gtgacttgatcagcgggaataccagtcaacatagtggtgaaatcaccgta 150
gttgaaaacagcttcagcaatttcaactgggtaagtttcagttggatgag 200
cagcttggaacatatagtattcagccaaatgagctctgatatctgagacg 250
tagacacctaattcgaccaggttaactctttcgtcagagggagataaagt 300
agtggtggctggggcagcagcgacaccagcagcaatagcagcgacaccag 350
caacaattgaagttagtttgaccatttttttcgattgaacttttgtagat 400
ctttttagtgaagatgtgagctcactcgaatgtaaataacaatgccaaat 450
tgtcggaaagagttaatcaaagctgctctatttatatgccgtttttttaat 500
aagcgacggacgaacagataaattgttgaatagctatttcactgctgata 550
tttctcttacttgggctcccctatcccatactcttcaccactacaaatat 600
gcagttgccctttcttcaacaatgctttttttatagatctcgtatacgga 650
tccgcgcctttgtactacctatatcttattatgatatatacaggagcaca 700
ggaatgttcggtacagggatgataccttaaaggaagttttggcatgcct 750
tgacaacttcaattaatctttggccaagaaaatgaaccagaaatcaaatt 800
ttattctgtgccctctgaacgagggcaatatccaatgtttgacactaaac 850
ggttgtcaggagaaaaattgaatgtttcccaaatcagaaacattaaaatc 900
cctctatatgatcagaggagtcgtacctgttagggtatgagcgaggaaac 950

FIG. 11

DR7

nt: SEQ ID NO: 7

```
aatgagttaccgtctgttacttttgggacggttttgcactaagaacaga      50
cgagtttacggttatcctcaacaagcaagcaagtatttgctaatctagat     100
gccattccgaatcattactcatacgttactattgagagatgttttacaat     150
agatgagaagaatacaatgtccagagctcctggtatgctagagtgcatat     200
tccaggtcttattcgaatcatatcataccgtccatttcaacaatggtgaa     250
atgtggtccacatatatcagaaatcttaacatttagtgaggagagccagt     300
agaaaaatgtgcgcaagcggaaagaagtcattcacagacacgtttaacaa     350
aacaccaccacagcagctttgtctcttgattctgatcagtttgccatcga     400
agaagcaaaattgtggtgttatttttttcaaacaaaacttttttggcaac     450
agcagttttcttctggatatttgtactttatcatccaaccgatgaaagct     500
ggtttcctgtcaacctacatttaaatggcccgtacttcttcaaaaccgct     550
agataagcaaattaacccaacttttgagcgtcctaaattccccttggctc     600
agaagactcgttaatatgggaagtttaagtcctaccatataatcaaattg     650
gaagctttctgtgttcgaatggctattctaaccgctgggctattaatcag     700
aggggaagtgaaatgaccgagacgtattatacgtcatgttgacatcaaca     750
atttaaggaaaaaaaataaaaaaaagcaatgaaaaagggtttttttaagtt     800
gaagacccttttcaaatatatgttgctttgaattgtatctaccgtctcgt     850
ttcttctgctttaccgtttttttttgccttctttagatatgtcttttatg     900
cttgaaaggtccggctttaatgcattcatctaaacgtagtattcctattt     950
ttgaactgctaccaatccaccatgactttact                       982
```

FIG. 12

DR8

nt: SEQ ID NO: 8

gtcaaactcccacagccaagtccaagagtacgaaaaaaaactttcacaac 50
gagttcaaaaaatgtgggatgaagactcccgtttttcaccgcctagaaaa 100
cagcgtttgctgagaaaaaaaataaatcatcgagaagaagtatgtcatca 150
taggatgttcccattgtaaggtgatgtgtaacatactcgaacaaagaatg 200
tatagagctgaatatttctcctttaaatttcaaagaaaatgagaaggaaa 250
atctcaaacagaaacttcgttctttttctcaagtaagcaaaagcttattg 300
agacaaagcggaataactacgatattaataacgttgatgaagctcgaaca 350
aagttagcgtcggttatgcttgcctatataaagatatatttgccttacat 400
tttcgttgaacgtagaatgatttttgcttttaataaatttttttgttgttc 450
tttcagtgcttcttcaactttgatacgaaagcaagtgcattagtacaaca 500
agaactggccacaactatactatactcatttttttcttgcccgtgttttaa 550
atgttttcatccacagcatttgatgggatgattggaagtgagacgttcga 600
gaaaatccatattttgagtcaagaattcagataatatactgagatgatta 650
ggtatggctgggttctacaaaaacacaaatatccggctagcaatgatcac 700
tgagcaaattaaagcgttaactcactcattattgtagcttatgcgtttct 750
cctcctctctttttttcctcgaaccggagtggaagatccaataacgtaat 800
attactgatgttgttattaaagctggcaaaaataacatgaggcgtaaaac 850
cgcactgcggtaagatgagggtataaggtggagatcaggcgaacaagctg 900
ttcta 905

FIG. 13

DR9

nt: SEQ ID NO: 9

```
aacgatccaatgagcgtttcatgatgccattgtttaatcagagtgatgaa      50
aaagaaatatttgcgacctttttcgttacattgatcgtgaaattttaat     100
caaagataatataaggacgtgagatatttatctttttacttgaaattaac    150
aatagaattgcgctaagcggaataagagctttcgtaaacctttctatttg    200
caccattgcgtcaacgtataaaatggtatgacctttacacaaacgcatgc    250
ttataatcttatgtttttcatagggtgtaatttggttgatgacgtagtct    300
aaatttgatgctatctgcaattgaggtacatataagaggtcaatttcggg    350
accaaccctttaatcgaaaaaaacgtaattcactagggcaagggagaac    400
ttagcagctaatatcgtaaacctttcatactaaaaaaatgcacttaccat    450
caacaaaaaactcaggaccaatttccaagcttttctaggtgattgcctat    500
aacacaaaaagattcgctcatacatgagatttttacatgtaatagcaatt    550
tgttccgatcagttgaaggtcatcaacgcacggcaggtacatccacacct    600
atcacaaagcccttcaataattcacctacgtaaagttataccgaaacatg    650
caaaatccatgaaaaattctgtatgataacgatcatatccttttgtattg    700
gtggtacgatgctcaaagatagttattgttgcacctgaggcaaaagcgga    750
aatgaaaaatccagatggggccaaaagcagaagtattgtgtacaacaatt    800
gcttcagcagtttaccaaaccgtttcccagcaatcatcaaaagttgcttt    850
agccacatttccgcaagatatctttgtggctcaacgaagagggctattcc    900
aaatgcaa                                              908
```

FIG. 14

DR10

nt: SEQ ID NO: 10

```
aactttcccccgtttaccacattgaagctgggtgtggaagatttatttga      50
agaaactaaaacgtaccctgtcatttcctgagtccccttcaacttagtg      100
tgaaagccgaacaattataatcctcggtagacaacagatttattgtacta      150
aagttactcttcctgttatcttccttgattttactgttatagcaatgacc      200
caccgcaatcaggagagccgccgtatggaatagcataccaagtcataaaa      250
tcgtcaacctattaacggggttcaggttctttttcagcgtagtagccctt      300
taacaagcgctgacaaagttgacactcagagaaaattcaggatttattgt      350
aatccagctactcatccttagatccgcttgcaggcatggtttttttcacc      400
ttgagaggctattttgggtaagccaggaaggctgaaaaatcccaaaagga      450
cacagtaataagaaattgttgttgttgtatgatgcatttagaactcaaaa      500
gacgagtttctgaaaatgcttacaatactccataggtaacatgatttttt      550
tattaaaaaagtatactgttcctttgggtaaaaattatgcaacccttgag      600
tgtccgatgaagataagactacgaaacaatttgcggtaaatttttttctgc      650
tattgacatttacacatgctccaatccattacccttttccattctcgtaat      700
aaaacctcgaactgttatttcatatttacatctagacgggtatcggcctc      750
aacaactccaaacaaaagtaaatagaaaagagccagacctatcgcaccgg      800
gtagagccagaaaatattttaaactatagttgacgtattctacggctgtt      850
gtttaggacaatacttttttccttcacaggcttcgaattacgcacatgcag      900
aactcctgt                                               909
```

FIG. 15

RC1

nt: SEQ ID NO: 11

```
gtgagacctccggaattttgacgctgcaagtcaatctacgggaaagaaga      50
aatttttaaacctaatgcaaaataagcttttcttggaaaataagatttt     100
cggcaataaaaggtaaatgcagccaaaaatcaaaatacttcagaagaagt     150
cgtagcgaggactgctaaggggaagcggatttgaagatcctttccagaac     200
aagaaggagccgaaagctgtcaggaactgttcctgatttttaggaaaac     250
aattaataggtatctcgtctagagtagtatctcgagcttccagaagttgc     300
agataatcaaaatcattgttttatccctttttttagattacagcttagaa     350
gagtagagagcaagtttactgaaacggttccttgtttacaataatattcc     400
taacaaactttacgaattaggatgcagcatgatttttttatattgcttcac     450
ttcctaaagtatgaatttttatccgtagtcgcaaacaaaacagctactgg     500
aaatctgcagcttgttaaaaaccggtagtttccgaatactcctcgtcctt     550
gagttgtataccgttaaacttcctagggtgtcatgtgtctggcccaattg     600
gcccacaaaatctggtcctattgacggttttcttttgattttcagcatct     650
tcctctaagaaggacagaaaattatgtaatatatgggagaaacggcctcc     700
caactgctaagtgtccccggcagcacgagtaagcaaaattcaggcaaact     750
attgcattaagaagccgtacataattcagcgtgatatgatgaaattttgt     800
taattgcaaattttagtacgatttggttgttagtgtgtgtttatgcaagt     850
aattattgaaccctaagtagttactgtcttcttttgctgtaattcgtgga     900
ttcacggccctccagcaacatggattgaa                          929
```

FIG. 16

RC2

nt: SEQ ID NO: 12

```
agtagcatttatgaccaaaagcgtacttaaattagcagcaaaaaaatttt     50
taataacgaaactataaggaaaatacgaggtactgattatgagagtcccc    100
gtttctcatttttgagacatgatctgaacaaggctgaaaacagcaatctt    150
tttcgataacttttgcaaaaatttcaaacattgttgtttgaatgcagcca    200
atttttatagggtacagagcttaatgctttacatgtgctttattttcggt    250
actttccttaaagtgtctacattatctctcaggacttgaatgtcttcggc    300
tgaattactataaaatcttgagttttctctgaagtttaatcctaagacaa    350
tagtggtgagtgatgtagttcacgtgtgtgccactggtaataatagagat    400
aactatctcagttaagtttgaaaaggtaaaaaatagtttaagtagtcatt    450
ttttgcgacggtcattcttctctgatgcacgttctttagactacctataa    500
acaccattcttacggaattataatggaaataaaacatcagtacgtgttgc    550
tgtcggtgatagaggggtaacagaaccttaattgaaaaattagcacagtg    600
cataatttattaacatgattgttttctgtgggaaataagaaatttcagca    650
ccagtaaaagacgagaaatatagggcacataaatgcgctcttactcgtat    700
gttccaggatgaaaatgtttagggcatcaagtattgccgaaagggcaata    750
tgctttaacaccagaaaatccactgtatactcgttacgggtaaacaaagc    800
aaaacgcagtgcgtgataatgtttctaaaatctctgcacactgttgaaat    850
gcggctctgatactttagcccttagtacctgacggtgcctaaaatgagga    900
```

FIG. 17

RC3

nt: SEQ ID NO: 13

```
tagttggaggttggtgagtaccagattgcttacaaaagaatagcgagcca      50
acatttgctctgcctcaggcctcttggtgctgcttgaagactcatcttat     100
atggcttttgtatgtcatgatttgttcttgtacattatgtgttgatatta     150
aacaaattgatttttttttttttgcgatagcaagcagataatgaaagaga     200
caaggacttggaacatccgataagactgcgccgatatcgatcttacagtc     250
cttcccttgtgtcatgactttcggaaaagcatcctcgtcgactggtagtt     300
tgctgtctgtcacgtgctgaagggtctgatacatttttttaaagataaga     350
gacggggtttacccttcggaggactaagcgagatctccaagtaaagatct     400
cgcttatcaagaaagcagccaagtgtggaacgtccttttttttggtttca     450
aaaagatattcaacagtttacactgcagctttaattgcctcaaaaggata     500
tcatgaggtgatctagggtcagaagggaaagattacagcatcttgagttg     550
aatcacatctgcaaaaggtggtattattgacgttgctcttccttaatgga     600
aactcatggggtttggaaaggaggtgcggtaatctatttttttcgaacac     650
aaaacctaaccttgaaaagaaactgtccaatttcattgaacttacctcag     700
aacgggccggagtctttgctttcagtctaacatggtctaatttcttcgaa     750
aagcttcatttaattgttagactgtggttttacaaggaaaaaaccagtgc     800
tatactgaagcgatacccagaactaattaccttgtgtgacgattcggctc     850
agcgaaacggacatggtaaaattgggaatttgaaagcaggcagcagcctt     900
gtacagcgacatgacgataggtttagaatccccatcacgtacgagttgaa     950
g                                                      951
```

FIG. 18

RC4

nt: SEQ ID NO: 14

```
tcctagagtagcgattccccttcgcgtattcttacatcttcgaagagaac      50
ttctggtgtaagtataataaatattatagctctatcgaatggtgcaatta     100
tttaccaaattctcaataggaatccataatactacatacgatactaatat     150
tctagtatttttatacttattatttctttttttattacaccagcaatcgtt     200
gcaaattatcttctgatagaatttctgagggtatcctaaacttatgccat     250
tttcttggactgtaaatcatacttggatgttgtgcattagtcaataatcg     300
gttcttgttccaacgattacatgtaaatgaagggagaaataattatggta     350
aatcatgcggcggtccttttggtgatgcagtatccatagtcactacataa     400
caatcttagtcaccttgtattgattcaccacataatcctgcagagcccgc     450
tatgtccttaatctgcgcgataactctcctacccctgaattttgagagcg     500
ccatagcaaaccgataaagctggcacaattaaaggtatcggtgttgtcag     550
aattaggtgcctcctgctttttttttttttcctgctcttatatccgttata     600
tccgaatgatttttatcgcttgtttaaaaaatactttcccgatatatata     650
tatagtctccctttaaatttgtttccggtaagtttttaacaccaataaat     700
gaaaagaaatgactacggtgatgaatatgagccgcgcattgaatcaggtt     750
atgtaagtatcagaacccctaattatgatgtcactcttacccttcgatgg     800
ctaagcggcgactgggatgccgggaaaagctctacaaatctactaaaaaa     850
gtcaaatatacagctgtaaacttctttcctcgtctacatcatggtaacga     900
ttgttcaatctttacttcgtgtctttttttttttctatgtactttctatt     950
ccaacctatgtgaagactaaaattcaccttagtaaacgtaaagacaatga    1000
cgataggtgc                                            1010
```

FIG. 19

RC5

nt: SEQ ID NO: 15

```
acaattgcatctgtgcttggctctcaagaacggtgtttggtgcatcaaaa      50
gttttcgactgcttatttggtcggaaatataaaaactcgatcctcttatc     100
taagcagtatacattcttctttttgaaatgaatgtactccgtaatatctt     150
cttatttggcattttcatccttaacttttgcatggctctgaactagtcag     200
atagttgccctttttcagcaaacctcttattattgaaagcatggtgtacat     250
ccgttatactattatattataagaaattgggatgccaatttttttgcttt     300
tgttttgcctgttttccttcttttcgcaaaagtaattgcagatttaatag     350
caggatattataccgttggtaaaacttaaggattttatgaacaatagctt     400
caagtacagcattcatagaaccaactactaaggatgaaactagtatgttt     450
ttgtcaaaatattttcttgaccttgctgtaacatcaagatctgtttctct     500
aagatattaaagttgagtaaaaacaaagctgatatgagaaaaatacgtaa     550
ttgctccacataatacgtgggtcagacataaaggtagaatacttgataca     600
gaagagattattcggtactcttgatggcgtgcttgaactggtgcctctta     650
acaaccggtaatatagtcagatgagtcactacgagtgtgtgtagtagcaa     700
gtgttttacctacgtggcagtaagagtagctctatggttgtgtaatagtg     750
gtgcttattcctaatgctctgaagtctgaagcggtacagttggtctggtc     800
tatatcatggtcaaaggagcaaacatatcttctgaagtgaccgcaaatag     850
tactatgatgtggttggcaatataacttaaaaggaaataaccacaaggaa     900
ttgcacccatgtacacagtttttcccggaaattgggaaaccagta         945
```

FIG. 20

RC6

nt: SEQ ID NO: 16

gcataatgcccgctataaaccttatttttatatgggggtctggcgcttc 50
gggaaaagagaggaaaacttgtaactcaatatatctcgatacaacattac 100
gttttgtaaatttatcacaaaagccaaatgatgatatctctcttgcaagt 150
tatcgaacattgattggtaatttgtttgaaaattgttaatttattgaata 200
tttcttttgcaaaagaaatagtctcagcgaaagctggttacaaaatttac 250
atcatgagtttacgggatttgtaaatacgcttttgcataaaaatacttt 300
gccgtttcccaccttgcatattcacttactcccccttcatatactcta 350
tgtaatgatgattaagctttggccgctaagtctctcaattagtgttgatt 400
ttggttttattcatatgattcttctttagtgaagtattgatcaattacgt 450
gagtcagcttttgaaaaccccatttggaaggaattaggaaattattttg 500
cttactacgaccactaatttaccgccatttctgggccttttttattgacta 550
ttttgaccatgtgctcgactagaagaacggcatcataatctgctggtaga 600
gttagtctataatgattgttgaaaataaaggcataagagatattccacct 650
aaaattcaagttattgactttattatcaggatcttagtatccttttttgg 700
taagtcatattcaatgaactaggtctcgcaaacttttttgttcgaaaagcg 750
gtagtgcatagttatgctaactctggatatatggcataaaccgtacaaca 800
ctagcccatttttttggaagtagtgagggcagctagactgtatgatgaat 850
attcgcctgcatactgagtttttttggtccttttttttatgtggctggcct 900
tacgatatg 909

FIG. 21

RC7

nt: SEQ ID NO: 17

```
gtttagtgcacccagactcgaatcttaaataccactttacacacctacta     50
aattttgtcctcacaaaatgaagacaggattcaaaaccgattaatagtag    100
cagaaactaaaaaagtacgaatattagtaaaattcatgttcttgaatcga    150
gctactatctttgtcgggagggtaaacgattataactcaaaatgactgga    200
actggtgattattaatttttacgtttcctgtgccaataagcggaagataa    250
gaggatagaagaaaagaaaggcggcacttggcgaactacaatggcgatta    300
tattcatggcgattatattcatacaaaggtaatggaggcctcggataatg    350
gacaatattgagaaaatccttatgcttacttctcttaataaaaaatagac    400
acagccatttattatgcgtaaaaaagattacccacttgtcttcgatgcgt    450
gctgctgccaatcaacctttgagcggaacttcgagctcgcaatgcgtct    500
ggaatgttgctagagacagtcttggttatctgtgacatgtgtttcgttca    550
ggcgtgtgagcatcttcttgttcgatttcaaaattaccgccttgactcgt    600
gaaactggataattcgttggcgttttcatataagtcgtctgatggcgaaa    650
acttttcctttacttagcatacagcaaatatccccatttgacggattttt    700
gaaaaatgagcccgctaacccagaatgaactgcattaccaagcatttatg    750
taaacgttccgccaccatctttggtaaggtatactattatgttctggatt    800
taaggttgattcacaatttttcatcaccaaaatctggtggcatgcctagt    850
tgtctggtttcaggcaatttagccatcatagaaaagcatcctctgtcttg    900
agttgagaaaatgttactcatagagccaaacaaataaaccctgg         944
```

FIG. 22

RC8

nt: SEQ ID NO: 18

```
ttcttgccgttctcatttcctgcacagtttctttgattatgtttgcagaa     50
gaatttctttatcgtttagtctaaacaaagaattcgttgtaaagaatttg    100
agagcggatcttgcatttttttatttatcatgcttatgtttttctttgat    150
gtaagaagaagcaagtaagatatgtgaatatcttatcactaattcaaata    200
actaagagagctcacaacgacaatttgtgacagcatgcgaagcaaagagc    250
agtgataccagtatctttcatccagtaataacatacgactgatgttatag    300
ttaaatgttacattttgagagacttcaacctctcgaaaccaagaggttgg    350
ttttaactctggtgacttcaagaagggtgggtacctttacaaagcttga    400
gacgaagcaatagtcagtctctgtataacaaggagaccacctcattttcc    450
agtaactcttgaggcatgtcggatggtttgccttgaataaaccgcagtca    500
ttataatgaatggcctgtactttcaaaacagtctggaaacagaaatccat    550
tgctgaggtacctttagtagcactttcgttagtgaaggtttaaggttag    600
ttcttatttactgcacaagagtttacatttaaccactctaatagtaactg    650
ttagagtggtttaactgttaggtgatctgttcattccatttttcgtgttg    700
tatctcaagatgagatagcttagcgttgctacatacataaatctaaacat    750
ataaacacctgtgtaactcgttaacgtctgggcttccatgcttctaccat    800
ttagaatgatgtagaccatttattccaagaggataagcaccctctgtgat    850
tcaaaatgataataagtgttgacgacaagttactctcgcagaattgttgt    900
caa                                                   903
```

FIG. 23

Utility1

nt: SEQ ID NO: 19

```
tgcaatttaaagagcgtacctgtaaataagaaggaagaacgttatgttat     50
taatggacttttagtgtcatcgaattttatgtaatatataagaaggtag     100
aataatttggcaggataatgtgttagcaaaggaggaaatcgaataccttt     150
aaaagagaaaaaatttttttagctgcttaaatttctgtgttataccacccg     200
atagattttgagttatgctttctaattgatctgactgcgaacgttttctt     250
tatgccatctgaattgtcaggaacaaagaagaaaaagaaaagtttttaaa     300
aaatctgtggtcgtgtgtgatgtacctttcctttacatgcattaatgcgc     350
tctgaaatgtggtacgatatccttacagagaatatattttctgtatatcg     400
tgcaatgttgaataacctatgaaggaaagtacccatcgctcaaggtaagc     450
attccaggagggtcgccagaaacttaaactagttttagcgacagatccga     500
aaattgatagagacattgaaaaaatcactactccgtccttttttagtgctt     550
tctcaatgcataatttggtgcacgactaaaaaattctagaacactatag     600
ttgcatttttttgggccggaagaagaaaaacgcatgtaactttaatgtcaa     650
ataaagttttcacctagtaagcgcgatacaaaaaaaacacagaaatagcc     700
ataggaaagtgaattttgtcagccgactaaaattaaggttagcttacaaa     750
gcagcaaaaaatttgacatcgcacggtattccctgaaaaaggagcaggca     800
ggtgctgtatatttttttcggttcctgcctcttacatggcgtcggtgtat     850
cttaaatactaaagtgagctgactacccttttgagtgccctatgtgacct     900
ctgatctcgaaagtaaacaagagatacctaatttcacagccacttttgtt     950
gcggacactgacgggatgtgttg                                973
```

FIG. 24

Utility2

nt: SEQ ID NO: 20

caaaaataacagcaagaaaagcggaaagaccatcgcaaggtggaaaggat 50
tataatggcacagcaaagtcgcacagagcactacagtatagcatagagtg 100
ctaatgagttgataggcccaattttgattatgccttctttttcatacacg 150
acgccagaggacattattacattacagtagttcgccgctagatgacaaac 200
gacatccttaccgatatgagatgtgcaaagctacataatggcaacaagcg 250
ttatgaacagccttgtctttacgaccacagaaaagccgtattagagctct 300
tcagctgcaaaattttcttctaatatgatgcaaagccatcaaaaatcatg 350
catagttatgaaatacctgatgaaacgcttcgagttcgtgctcaagaaat 400
tactgaaaggttaccgagaagaaaaatatctatgagacacgataaggccc 450
cttctgaatccattgtcctgggcttgttcattctatttaccacttaaaat 500
tgatcctttcaaaggaatttttttctatttccaatagtatatttgtacaa 550
aaactacaaaaatggataaaaaataacagtaatttgtgactactgtaaat 600
atcactgatttggattttgtaatgagtactgctcatgcccatgccgatgc 650
aagtggatcataaattttactaaacgatattcgataatgcgccaagcctt 700
tataaggaactcaaaataacccatatggacagtttcagaaggccaaataa 750
cgatcaaggacattcactcatgtttttcaaaggcgaagagtgtaaaattt 800
tcttctatatagttcgaatattttatcttataaatttcagtcgtcatttt 850
ccacattcgaactcaaataatgataaagaacgctgcagtaatggcttaaa 900

FIG. 25

Utility3

nt: SEQ ID NO: 21

gcaagtatggggtagcaagctgcttaaagcttcttatcacactgtaacct 50
ctgttaaagcagctgcgttgttcttgcactctgcacgtaaatgacgacgg 100
ccaatgtaaaaatagcagtcactgcaggcctcttggattgtacccgtaat 150
tacagcaattgccattagatttacgaagtcatttaattaaatggtttgat 200
ttatctccttaaatgcttccagaaatgggactagacttttttcactcaaa 250
cctgttcacaattattgctctttttcaattataaggtaaacaaggccatc 300
tatcagcaacacagtgctcgcatttttttaattaaactatataaaaccaac 350
tatttgtggttgcgacttcacttttttgttgaattactacccaatcattaa 400
tattgaagatgtgagatcatagatttattggctttgggcatctcaaatcc 450
caagaggtcattttaaccaacaacattttaaaaagtagatttgtctgcct 500
cagctatgagatgcgcatgtccctagcatctcatatctggttatattatt 550
ttttccacttggtgaatgttgaaaaaaacaccactcgtccaatttatcag 600
tttgcaggtctaatgtccttccctgttatttaaactgtatattgtaagca 650
tgtcttatcgaaacaacttactcagttgtccgaaaacaaaactgcaaatt 700
ctgtgtgtattcacgtactagaatcctgtcaaattggatcttgatttaag 750
cttttatagcaacgaactttgcatactaagttttttttgttaaccggaac 800
tgccaagaagcattcagtaaaatacatcttcatcatttactgataatact 850
cattcagactcatatcatactatttcgaattcattatacatcctcaaaaa 900
ccatattcttcagttgtaataaaagatagagcctgcatttgattcgattt 950

FIG. 26

Negative1

nt: SEQ ID NO: 22

```
gatttaatacagtacctttcttcgctaggatctatatgcgaatatatcac      50
atatgtaaattataagctcatcgcaaaaccaaaaaaaaaaaaattttcaa     100
taatttttcactaatcttcaaaaacaaatggggtaacccgtacaagagtt     150
attaaaacccaaaatgacaaaatcgcgacaattcaatcctacttaattag     200
caataacatactagcggtagagctactatcacatgttgaaccttgaatgc     250
tcaattcattgtactcaatactgctatcaaaagaaaaaaaatgtattaat     300
tatattcttgtcaaaatcaattttacactataagaggaaaatgttcttca     350
gtcctagtaacattagttttctccctttgctagagactttacataatatc     400
ctagaaggtaaaattcgataatacagcagtaaagtcgtatattggtagca     450
atccttggtgacgctgacttttttttttgtaattttattgtttagttca     500
tgataaaaaacttcaaatcacttttaatctggtagacagagaaaacaaat     550
cgaaacgaaaatagagaactacgaataaaaaaatataagtggagaagatc     600
gtcactacgcattaaacaatattgatcgctcaatgccagtactgcgcgta     650
aaagtttagtaacttaacgatttaggcacaatttgagaaaaatttcgccc     700
tgcagtaagtatgttattcagtacgatataaagctgaggttttatgctgg     750
caacgttcagatttttaggttatcagcaatgttaaaatattaaatagga     800
tacttttattgtttgagaccaccctcaatgccagatatgttaaacgcttt     850
tttctggagtgaggtatcatagaaaaaggctcgagtacatcaagcactta     900
aaggttcaacactctactgttacttctttaagctaagctattcatacata     950
atagtccatcaaagtgg                                      967
```

FIG. 27

Negative2

nt: SEQ ID NO: 23

```
gcaatttgcagttcaacttttcaatgatgatttagaatgatctaacactg      50
gaagttgaagtttttcaaaaatttgctgtaaaatttgaccgatttgtgag     100
attcttcctggctgtcagaatatggggccgtagtatattgtcagacctgt     150
tcctttaagaggtgatggtgataggcgttgagtatgtgtagtgtttgacc     200
cgagggtatggttttcacaagtactgcgcactgtattgtgaaagcagctt     250
cggggtgcgtgattaaaaaatgcgaccaagaataaacaggtacatcataa     300
caagggccatttgaattgcatttatcaggatttgtaaccttgttctaaag     350
aggcatcgtatagtttaagttcattttccacccaatttgatgacggtgtg     400
gaccttaacctattgtcttgaaatttaggttatctcttagatatcacatg     450
tgattaccccagtgaacgcgtataagcttacagaaaggaaaaccggttgg     500
ctcagtcaaaactgttgcagatttgggctcccctgaatatttgagacatc     550
cctaaaatgaagagatatatacagctaattttgaatgaaaatttaaaatt     600
cgcaatgaacagtactagagatgagcttttgaagtcctttcaaattattt     650
gttcttccagttgatatttttttatttttatataccagtaccaaatataatc     700
ttgccatacatttacctttttgaggttgttcaacggaaatccagtgtatt     750
tacacattcttggaaacccatcgcttataatacgaactaatttatttatg     800
aacaaaggctttggaaaagtatccctacttttttacgacgctaaatcatga     850
tacgaaactttaggaagattaacagtcactccataaaatcagaaagtatt     900
cgctaatagtggaaagaaatggttatataaagatggaaatatcttgaaag     950
agacagtttaacccgaagttctgtcaaagtg                        981
```

FIG. 28

DR1s

nt: SEQ ID NO: 24

```
ttcagaaaagcaaggaaacagtactatcgtttagaatgtagaatgatagg      50
ttgcttgctaattctattatggcacgaatgatacacccatattttcaaca     100
aaatcaatacccactagcatcattgagccaactatttgtcaatgcaacca     150
ttaccggtacttcatcctgatttaacgagtctacttttttatcacgtcaa     200
aatttacttgttttcctgtaaacccgaaataaaggcaaaaaagacctggg     250
tgcaattacgaataaatgtacaataatcatcctgtttgcatagtaaactt     300
ccagttagagtcacacaacgcaatgaattttgacagttttctgtgcgata     350
ttctttggtaaacgtaaagaacaggcaacttttggtacaatggattctag     400
cccatatggttcatttctggtgcattcgcaaagtcagtatttgtctagct     450
gtgttttctggctgagagacattatgatgttattcattgttatggatatc     500
tctgtagctcatgctgcttatttctccctaaaaaagtttttttctctcgaa     550
tacattcttgaccatttcatagtgaaattcttgtacttatttaaaaccaa     600
aaatggaagtattcatacatccccctatcaaaaacactcaataagtttcg     650
aattattcgttcgtctaaacagtgtccaatactcaaaggggtattcaaga     700
cggcacaaaatcagcatcttcccttatccgtgttccagaaataccacgct     750
aaggtttttcctcctacaatccataaaatcattaaggaggcagcttgaaa     800
aatcttg                                                807
```

FIG. 29

DR2s

nt: SEQ ID NO: 25

```
tttctttttccctatttctcactggttgacagaaatcagtgtgctatcat      50
cctaccatatgcgctaaacttattgtctttctcctcctagagatgctgta     100
ttccatgcatattctgaacgatgggttggtgtttttatcaagcaaggtta     150
atcacatggcgtggcttgctccacacatcagtagaaaacgcataccgcag     200
cggaatccttaaataataagtgattttactgttcatcaactacaatcgga     250
ctctttcacaattacccttcttgttttccacatttactgttaaatgaagg     300
gatgtacagaaggcttaggaaaacctgtgctgaatactggatggacactg     350
cattcccacagtgaaacttttatagatacactgtcagttattttcgaact     400
ttcatcaagttgctgagttttagtatccctttgccttagctatatgtttg     450
aatgagcaaaatatttgcaatgtctctagctttcttgaaatattggttta     500
tattgagggcttggtaagatttcaaatttcactttgaaatactcaggaga     550
aaaatcatgctcttttgataatttggtgactaaacatacataaaacagtt     600
taattttgggtggtaatggctgtgtgactagctatagaaagaaaaaaatt     650
aaaaaaaaaaaaaaaaatcaagtagttcctgcactgcgacgtccattata     700
gcattatgaattggtccctgatttacgcatgcgataaactatttttagcg     750
cagccgcatatt                                           762
```

FIG. 30

DR3s

nt: SEQ ID NO: 26

```
aaaatctcaaaattcccaatattcacatagtctaaagtaccgatagcaac      50
caacatatataaacagtagtattttacgaagctgaattgcaagattagtg     100
agaggagaataaccggataatttttttggattacgttattgttaaaggct     150
ataatattaggtgaaacagaatgtcctagaagtttttttctttcatgtta     200
aatttattgattcttgcgcttcagcttttataaaacataagaactgtttc     250
ttcacgttaacttcttgtgccacatataatgatgtactagtaatatgggt     300
actatttggcagatgatatttgatttttattcaagacggttactgtttct     350
acgattgatattttcattcctggatatcatcttgccagatcacttacaat     400
ttaggccgcgcctgaattgaagagtacttcaatacgtagtgtactgtcca     450
aactctcttccaaatttttaatatttagctggggttgggtaacaagtgag     500
caagggaaaaagtgaacattttaagaagaacaataaaatagcaagagatg     550
gaatggtaatgcttggctctcgagaagagtagcataaaacgagacttgtt     600
taaaacaggatatgacatacttcaattcagctttccctatcagccgctcg     650
agcagttatataggtgtgttgccggagtaatttggcggaggccaacagtg     700
gctaggcggcaacgcctggaacacgcgcttaaaagttctggaaggttcgc     750
gaattgagaactgctcaggggcgaatacaggggcggccttggcggcaggg     800
gggaggcctctgtgaagttagttatataagacttgctgtcatcgtttttt     850
tgatcccggcaggaactatctttt                               874
```

FIG. 31

DR4s

nt: SEQ ID NO: 27

```
aatagatgaattatgtgccgctggacgtttatagatagcataagcacaat     50
gacttaaaggttataatactcattgatatcactctgattataaaatcgta    100
atatgcgaataggtgaactaatcggaataacccatacgacacttcaagct    150
tcaattctatttcaactgtagtgcctgctagtgaagaatacaaaagtagc    200
atacgtgatgtgcaaaaaatgcgctacttatcacacaagtaccttgcgca    250
agaagggtactctaaaccggggccatcgcattaccagacggagatgtatt    300
ctttatgaagcaataattggaggtgtatcaagttcgaaactgctgatgct    350
atggatttacatctttcttatgcacaaggcttgcttgtgtttctgagtag    400
ttagtttttagatttttgtcaagtctggggtaagttaattcgagcaaaat    450
taacggcacgttattctaatgcatatgttgttcatatattcttttacaaa    500
gaggtttggaatgatgtcaccgatgttagaatgttaggagaatttcatgt    550
gaattttagtccaagtgttgaagttctcttctgcagttagggcacgtaca    600
tggcaacgatatcgtttttgatgtattaatcttagtaggcgttgagtttg    650
tatgttacttttctcaggtgatgaagcgtgatgacgatgacaaaaatggg    700
ttataatagggcgcactatcatcatgcgtgattgatatttaaccaatgtc    750
ttgagtacatcaactccagaaaatgggtcattatatgcctagcatgt      797
```

FIG. 32

DR5s

nt: SEQ ID NO: 28

```
tggcaatgatactcgttattcgtaatatcagtccgtcaaggtgctgtgat      50
ttctctattttatattgcctattatttttcaaatgatttgagccgtttt     100
aaattgagtatgcaatgagtcttttgaatcaaccgtaaggcagttccata    150
accactgccacgaatacgtttcactaccttgaagaatctctaatgtaggc    200
cgtattcttcgcacttagttctgacgatgtagacatctcattatataaga    250
gcataagcgcctgtttctagaatcatttcttcgtgacccagcttttgag     300
ttatttcgcggtattttgaaacatttctcgagcttgacgtgaacatcctt    350
atatttcatgacaaactcgatcattggaacatccctgcctcgattttaga    400
gctagtatcaaatttcaatctctttgtgatggagccccgctcctatttca    450
aaagagaagtttcttgtatgcatatgttattgaagtctgattatagcaag    500
tgcaatgtcgtctcaattattttaactatttttagccatacatgttagtt    550
atcctcaaagagagcctccagactgggaagcagtgtttgtcatttcaaat    600
aagtagatttcacagtttgtatgattttcgaagccaggattcattgggct    650
ttgagtaaagagaagccgcgtattacgaacagcttacgatattgtaaaat    700
attcccttattgtggtgccccaatggatacatgccagagaaatgtctgtg    750
aaattgaacaattacaatgacgagagcaagtaatccggcggccttgtctc    800
tctttcac                                              808
```

FIG. 33

DR6s

nt: SEQ ID NO: 29

agataccgtccttggatagagcgctggagatagctggtctcaatctggtg 50
gagtaccatgggacaccagtgatgactctagtgacttgatcagcgggaat 100
accagtcaacatagtggtgaaatcaccgtagttgaaaacagcttcagcaa 150
tttcaactgggtaagtttcagttggatgagcagcttggaacatatagtat 200
tcagccaaatgagctctgatatctgagacgtagacacctaattcgaccag 250
gttaactctttcgtcagagggagataaagtagtggtggctggggcagcag 300
cgacaccagcagcaatagcagcgacaccagcaacaattgaagttagtttg 350
accatttttttcgattgaacttttgtagatcttttttagtgaagatgtgag 400
ctcactcgaatgtaaataacaatgccaaattgtcggaaagagttaatcaa 450
agctgctctatttatatgccgtttttttaataagcgacggacgaacagata 500
aattgttgaatagctatttcactgctgatatttctcttacttgggctccc 550
ctatcccatactcttcaccactacaaatatgcagttgccctttcttcaac 600
aatgctttttttatagatctcgtatacggatccgcgcctttgtactacct 650
atatcttattatgatatatacaggagcacaggaatgttcggtacagggat 700
gataccttt 709

FIG. 34

DR7s

nt: SEQ ID NO: 30

```
ttgggacggtttttgcactaagaacagacgagtttacggttatcctcaac      50
aagcaagcaagtatttgctaatctagatgccattccgaatcattactcat     100
acgttactattgagagatgttttacaatagatgagaagaatacaatgtcc     150
agagctcctggtatgctagagtgcatattccaggtcttattcgaatcata     200
tcataccgtccatttcaacaatggtgaaatgtggtccacatatatcagaa     250
atcttaacatttagtgaggagagccagtagaaaaatgtgcgcaagcggaa     300
agaagtcattcacagacacgtttaacaaaacaccaccacagcagctttgt     350
ctcttgattctgatcagtttgccatcgaagaagcaaaattgtggtgttat     400
ttttttcaaacaaaacttttttggcaacagcagttttcttctggatattt     450
gtactttatcatccaaccgatgaaagctggtttcctgtcaacctacattt     500
aaatggcccgtacttcttcaaaaccgctagataagcaaattaacccaact     550
tttgagcgtcctaaattccccttggctcagaagactcgttaatatgggaa     600
gtttaagtcctaccatataatcaaattggaagctttctgtgttcgaatgg     650
ctattctaaccgctgggctattaatcagaggggaagtgaaatgaccgaga     700
cgtattatacgtcatgttgacatcaacaatttaaggaaaaaaataaaaaa     750
aagcaatgaaaaagggtttttttaagttgaagacccttttcaaatatatg     800
ttgctttgaattgtatctaccgtctcgtttcttctgctttaccgtttttt     850
tttgccttctttagatatgtcttttatgcttgaaaggtccggc           893
```

FIG. 35

DR8s

nt: SEQ ID NO: 31

```
ttcaccgcctagaaaacagcgtttgctgagaaaaaaaataaatcatcgag      50
aagaagtatgtcatcataggatgttcccattgtaaggtgatgtgtaacat     100
actcgaacaaagaatgtatagagctgaatatttctcctttaaatttcaaa     150
gaaaatgagaaggaaaatctcaaacagaaacttcgttctttttctcaagt     200
aagcaaaagcttattgagacaaagcggaataactacgatattaataacgt     250
tgatgaagctcgaacaaagttagcgtcggttatgcttgcctatataaaga     300
tatatttgccttacattttcgttgaacgtagaatgatttttgcttttaat     350
aaatttttttgttgttctttcagtgcttcttcaactttgatacgaaagcaa     400
gtgcattagtacaacaagaactggccacaactatactatactcatttttt     450
cttgcccgtgttttaaatgttttcatccacagcatttgatgggatgattg     500
gaagtgagacgttcgagaaaatccatattttgagtcaagaattcagataa     550
tatactgagatgattaggtatggctgggttctacaaaaacacaaatatcc     600
ggctagcaatgatcactgagcaaattaaagcgttaactcactcattattg     650
tagcttatgcgtttctcctcctctctttttttcctcgaaccggagtggaa     700
gatccaataacgtaatattactgatgttgttattaaagctggcaaaaata     750
acatgaggcgtaaaaccgcactgcggtaagatgagggt                 788
```

FIG. 36

DR9s

nt: SEQ ID NO: 32

```
tgaaaaagaaatatttgcgacctttttcgttacattgatcgtgaaattt     50
taatcaaagataatataaggacgtgagatatttatctttttacttgaaat    100
taacaatagaattgcgctaagcggaataagagctttcgtaaacctttcta    150
tttgcaccattgcgtcaacgtataaaatggtatgacctttacacaaacgc    200
atgcttataatcttatgtttttcatagggtgtaatttggttgatgacgta    250
gtctaaatttgatgctatctgcaattgaggtacatataagaggtcaattt    300
cgggaccaaccctttttaatcgaaaaaaacgtaattcactagggcaaggga  350
gaacttagcagctaatatcgtaaacctttcatactaaaaaaatgcactta    400
ccatcaacaaaaaactcaggaccaatttccaagcttttctaggtgattgc    450
ctataacacaaaaagattcgctcatacatgagattttacatgtaatagc     500
aatttgttccgatcagttgaaggtcatcaacgcacggcaggtacatccac    550
acctatcacaaagcccttcaataattcacctacgtaaagttataccgaaa    600
catgcaaaatccatgaaaaattctgtatgataacgatcatatccttttgt    650
attggtggtacgatgctcaaagatagttattgttgcacctgaggcaaaag    700
cggaaatgaaaaatccagatggggccaaaagcagaagtattgtgtacaac    750
aattgcttcagcagtttaccaaaccgtttcccagcaatcatcaaaagttg    800
ctttagccacatttccgcaagatat                             825
```

FIG. 37

DR10s

nt: SEQ ID NO: 33

```
attgaagctgggtgtggaagatttatttgaagaaactaaaacgtaccctg      50
tcatttcctgagtccccttttcaacttagtgtgaaagccgaacaattataa     100
tcctcggtagacaacagatttattgtactaaagttactcttcctgttatc      150
ttccttgattttactgttatagcaatgacccaccgcaatcaggagagccg      200
ccgtatggaatagcataccaagtcataaaatcgtcaacctattaacgggg      250
ttcaggttctttttcagcgtagtagccctttaacaagcgctgacaaagtt      300
gacactcagagaaaattcaggatttattgtaatccagctactcatcctta      350
gatccgcttgcaggcatggtttttttcaccttgagaggctattttgggta      400
agccaggaaggctgaaaaatcccaaaaggacacagtaataagaaattgtt      450
gttgttgtatgatgcatttagaactcaaaagacgagtttctgaaaatgct      500
tacaatactccataggtaacatgatttttttattaaaaaagtatactgtt      550
cctttgggtaaaaattatgcaacccttgagtgtccgatgaagataagact      600
acgaaacaatttgcggtaaatttttttctgctattgacatttacacatgct      650
ccaatccattacccttttccattctcgtaataaaacctcgaactgttattt      700
catatttacatctagacgggtatcggcctcaacaactccaaacaaaagta      750
aatagaaaagagccagacctatcgc                               775
```

FIG. 38

RC1s

nt: SEQ ID NO: 34

```
gctgcaagtcaatctacgggaaagaagaaatttttaaacctaatgcaaa      50
ataagcttttcttggaaaataagattttcggcaataaaaggtaaatgcag     100
ccaaaaatcaaaatacttcagaagaagtcgtagcgaggactgctaagggg     150
aagcggatttgaagatcctttccagaacaagaaggagccgaaagctgtca     200
ggaactgttcctgatttttaggaaaacaattaataggtatctcgtctag     250
agtagtatctcgagcttccagaagttgcagataatcaaaatcattgtttt     300
atccctttttttagattacagcttagaagagtagagagcaagtttactga     350
aacggttccttgtttacaataatattcctaacaaactttacgaattagga     400
tgcagcatgatttttatattgcttcacttcctaaagtatgaatttttat      450
ccgtagtcgcaaacaaaacagctactggaaatctgcagcttgttaaaaac     500
cggtagtttccgaatactcctcgtccttgagttgtataccgttaaacttc     550
ctagggtgtcatgtgtctggcccaattggcccacaaaatctggtcctatt     600
gacggttttcttttgattttcagcatcttcctctaagaaggacagaaaat     650
tatgtaatatatgggagaaacggcctcccaactgctaagtgtccccggca     700
gcacgagtaagcaaaattcaggcaaactattgcattaagaagccgtacat     750
aattcagcgtgatatgatgaaattttgttaattgcaaattttagtacgat     800
ttggttgttagtgtgtgtttatgcaagtaattattgaaccctaagtagtt     850
actgtcttcttttgctgtaattcgtggattcacg                    884
```

FIG. 39

RC2s

nt: SEQ ID NO: 35

```
gtccccgtttctcatttttgagacatgatctgaacaaggctgaaaacagc      50
aatctttttcgataacttttgcaaaaatttcaaacattgttgtttgaatg     100
cagccaatttttatagggtacagagcttaatgctttacatgtgctttatt     150
ttcggtactttccttaaagtgtctacattatctctcaggacttgaatgtc     200
ttcggctgaattactataaaatcttgagttttctctgaagtttaatccta     250
agacaatagtggtgagtgatgtagttcacgtgtgtgccactggtaataat     300
agagataactatctcagttaagtttgaaaaggtaaaaaatagtttaagta     350
gtcatttttgcgacggtcattcttctctgatgcacgttctttagactac     400
ctataaacaccattcttacggaattataatggaaataaaacatcagtacg     450
tgttgctgtcggtgatagaggggtaacagaaccttaattgaaaaattagc     500
acagtgcataatttattaacatgattgttttctgtgggaaataagaaatt     550
tcagcaccagtaaaagacgagaaatatagggcacataaatgcgctcttac     600
tcgtatgttccaggatgaaaatgtttagggcatcaagtattgccgaaagg     650
gcaatatgctttaacaccagaaaatccactgtatactcgttacgggtaaa     700
caaagcaaaacgcagtgcgtgataatgtttctaaaatctctgcacactgt     750
tgaaatgcggctctgatactttagcc                             776
```

FIG. 40

RC3s

nt: SEQ ID NO: 36

```
ccagattgcttacaaaagaatagcgagccaacatttgctctgcctcaggc      50
ctcttggtgctgcttgaagactcatcttatatggcttttgtatgtcatga     100
tttgttcttgtacattatgtgttgatattaaacaaattgatttttttttt     150
tttgcgatagcaagcagataatgaaagagacaaggacttggaacatccga     200
taagactgcgccgatatcgatcttacagtccttcccttgtgtcatgactt     250
tcggaaaagcatcctcgtcgactggtagtttgctgtctgtcacgtgctga     300
agggtctgatacatttttttaaagataagagacggggtttacccttcgga     350
ggactaagcgagatctccaagtaaagatctcgcttatcaagaaagcagcc     400
aagtgtggaacgtccttttttttggtttcaaaaagatattcaacagttta     450
cactgcagctttaattgcctcaaaaggatatcatgaggtgatctagggtc     500
agaagggaaagattacagcatcttgagttgaatcacatctgcaaaaggtg     550
gtattattgacgttgctcttccttaatggaaactcatggggtttggaaag     600
gaggtgcggtaatctatttttttcgaacacaaaacctaaccttgaaaaga     650
aactgtccaatttcattgaacttacctcagaacgggccggagtctttgct     700
ttcagtctaacatg                                         714
```

FIG. 41

RC4s

nt: SEQ ID NO: 37

```
ttcgcgtattcttacatcttcgaagagaacttctggtgtaagtataataa      50
atattatagctctatcgaatggtgcaattatttaccaaattctcaatagg     100
aatccataatactacatacgatactaatattctagtatttttatacttat     150
tatttctttttattacaccagcaatcgttgcaaattatcttctgataga     200
atttctgagggtatcctaaacttatgccattttcttggactgtaaatcat     250
acttggatgttgtgcattagtcaataatcggttcttgttccaacgattac     300
atgtaaatgaagggagaaataattatggtaaatcatgcggcggtcctttt     350
ggtgatgcagtatccatagtcactacataacaatcttagtcaccttgtat     400
tgattcaccacataatcctgcagagcccgctatgtccttaatctgcgcga     450
taactctcctacccctgaattttgagagcgccatagcaaaccgataaagc     500
tggcacaattaaaggtatcggtgttgtcagaattaggtgcctcctgcttt     550
ttttttttcctgctcttatatccgttatatccgaatgatttttatcgct     600
tgtttaaaaaatactttcccgatatatatatatagtctccctttaaattt     650
gtttccggtaagtttttaacaccaataaatgaaaagaaatgactacggtg     700
atgaatatgagccgcgcattgaatcaggttatgtaagtatcagaacccct     750
aattatg                                                757
```

FIG. 42

RC5s

nt: SEQ ID NO: 38

```
ctcaagaacggtgtttggtgcatcaaaagttttcgactgcttatttggtc     50
ggaaatataaaaactcgatcctcttatctaagcagtatacattcttcttt    100
ttgaaatgaatgtactccgtaatatcttcttatttggcattttcatcctt    150
aacttttgcatggctctgaactagtcagatagttgccctttcagcaaac    200
ctcttattattgaaagcatggtgtacatccgttatactattatattataa    250
gaaattgggatgccaatttttttgcttttgttttgcctgttttccttctt    300
ttcgcaaaagtaattgcagatttaatagcaggatattataccgttggtaa    350
aacttaaggattttatgaacaatagcttcaagtacagcattcatagaacc    400
aactactaaggatgaaactagtatgtttttgtcaaaatattttcttgacc    450
ttgctgtaacatcaagatctgtttctctaagatattaaagttgagtaaaa    500
acaaagctgatatgagaaaaatacgtaattgctccacataatacgtgggt    550
cagacataaaggtagaatacttgatacagaagagattattcggtactctt    600
gatggcgtgcttgaactggtgcctcttaacaaccggtaatatagtcagat    650
gagtcactacgagtgtgtgtagtagcaagtgttttacctacgtggcagta    700
agagtagctctatggttgtgtaatagtggtgcttattcctaatgctctga    750
agtctgaagcggtacagttggtctggtctatatcatggtcaaaggagcaa    800
acatatcttctgaagtgaccgcaaatagtactatgatgtggttggcaata    850
taacttaaaaggaaataaccacaaggaattgcacccatgta             891
```

FIG. 43

RC6s

nt: SEQ ID NO: 39

```
tttttatatgggggtctggcgcttcgggaaaagagaggaaaacttgtaac      50
tcaatatatctcgatacaacattacgttttgtaaatttatcacaaaagcc     100
aaatgatgatatctctcttgcaagttatcgaacattgattggtaatttgt     150
ttgaaaattgttaatttattgaatatttcttttgcaaaagaaatagtctc     200
agcgaaagctggttacaaaatttacatcatgagtttacgggatttgtaaa     250
tacgctttttgcataaaaatactttgccgtttcccaccttgcatattca     300
cttactccccccttcatatactctatgtaatgatgattaagctttggccg     350
ctaagtctctcaattagtgttgattttggttttattcatatgattcttct     400
ttagtgaagtattgatcaattacgtgagtcagcttttgaaaaccccatt     450
tggaaggaattaggaaattattttgcttactacgaccactaatttaccgc     500
catttctgggccttttttattgactattttgaccatgtgctcgactagaag     550
aacggcatcataatctgctggtagagttagtctataatgattgttgaaaa     600
taaaggcataagagatattccacctaaaattcaagttattgactttatta     650
tcaggatcttagtatccttttttggtaagtcatattcaatgaactaggtc     700
tcgcaaacttttttgttcgaaaagcggtagtgcatagttatgctaactctg     750
gatatatggcataaaccgtacaacactagcccatttttttggaagtagtg     800
agggcagctagactgtatgatgaatattcgcctgcatactgagttttt      848
```

FIG. 44

RC7s

nt: SEQ ID NO: 40

```
tgaagacaggattcaaaaccgattaatagtagcagaaactaaaaaagtac      50
gaatattagtaaaattcatgttcttgaatcgagctactatctttgtcggg     100
agggtaaacgattataactcaaaatgactggaactggtgattattaattt     150
ttacgtttcctgtgccaataagcggaagataagaggatagaagaaaagaa     200
aggcggcacttggcgaactacaatggcgattatattcatggcgattatat     250
tcatacaaaggtaatggaggcctcggataatggacaatattgagaaaatc     300
cttatgcttacttctcttaataaaaaatagacacagccatttattatgcg     350
taaaaaagattacccacttgtcttcgatgcgtgctgctgccaatcaacct     400
tttgagcggaacttcgagctcgcaatgcgtctggaatgttgctagagaca     450
gtcttggttatctgtgacatgtgtttcgttcaggcgtgtgagcatcttct     500
tgttcgatttcaaaattaccgccttgactcgtgaaactggataattcgtt     550
ggcgttttcatataagtcgtctgatggcgaaaacttttcctttacttagc     600
atacagcaaatatccccatttgacggatttttgaaaaatgagcccgctaa     650
cccagaatgaactgcattaccaagcatttatgtaaacgttccgccaccat     700
ctttggtaaggtatactattatgttctggatttaaggttgattcacaatt     750
tttcatcaccaaaatctggtggcatgcctagttgtctggtttcaggcaat     800
ttagcc                                                 806
```

FIG. 45

RC8s

nt: SEQ ID NO: 41

```
aatttgagagcggatcttgcatttttatttatcatgcttatgtttttc      50
tttgatgtaagaagaagcaagtaagatatgtgaatatcttatcactaatt     100
caaataactaagagagctcacaacgacaatttgtgacagcatgcgaagca     150
aagagcagtgataccagtatctttcatccagtaataacatacgactgatg     200
ttatagttaaatgttacattttgagagacttcaacctctcgaaaccaaga     250
ggttggttttaactctggtgacttcaagaagggtgggtacctttacaaa     300
gcttgagacgaagcaatagtcagtctctgtataacaaggagaccacctca     350
ttttccagtaactcttgaggcatgtcggatggtttgccttgaataaaccg     400
cagtcattataatgaatggcctgtactttcaaaacagtctggaaacagaa     450
atccattgctgaggtacctttagtagcactttcgttagtgaaggtttaa     500
ggttagttcttatttactgcacaagagtttacatttaaccactctaatag     550
taactgttagagtggtttaactgttaggtgatctgttcattccatttttc     600
gtgttgtatctcaagatgagatagcttagcgttgctacatacataaatct     650
aaacatataaacacctgtgtaactcgttaacgtctgggcttccatgcttc     700
taccatttagaatgatgtagaccatttattccaagaggataagcaccctc     750
tgtgattcaaaat                                          763
```

FIG. 46

Utility1s

nt: SEQ ID NO: 42

```
ttaatggacttttagtgtcatcgaattttatgtaatatataagaaggta      50
gaataatttggcaggataatgtgttagcaaaggaggaaatcgaatacctt     100
taaaagagaaaaaatttttttagctgcttaaatttctgtgttataccaccc    150
gatagattttgagttatgctttctaattgatctgactgcgaacgttttct     200
ttatgccatctgaattgtcaggaacaaagaagaaaaagaaaagtttttaa     250
aaaatctgtggtcgtgtgtgatgtacctttcctttacatgcattaatgcg     300
ctctgaaatgtggtacgatatccttacagagaatatattttctgtatatc     350
gtgcaatgttgaataacctatgaaggaaagtacccatcgctcaaggtaag     400
cattccaggagggtcgccagaaacttaaactagttttagcgacagatccg     450
aaaattgatagagacattgaaaaaatcactactccgtccttttttagtgct    500
ttctcaatgcataattttggtgcacgactaaaaaattctagaacactata     550
gttgcatttttttgggccggaagaagaaaaacgcatgtaactttaatgtca    600
aataaagttttcacctagtaagcgcgatacaaaaaaaacacagaaatagc     650
cataggaaagtgaattttgtcagccgactaaaattaaggttagcttacaa     700
agcagcaaaaaatttgacatcgcacggtattccctgaaaaaggagcaggc     750
aggtgctgtatatttttttcggttcctgcctcttacatggcgtcggtgta     800
tcttaaatactaaagtgagctgactacccttttgagtgccctatgtgacc     850
tctgatctcgaaagtaaacaaga                                873
```

FIG. 47

Utility2s

nt: SEQ ID NO: 43

```
ttataatggcacagcaaagtcgcacagagcactacagtatagcatagagt      50
gctaatgagttgataggcccaattttgattatgccttctttttcatacac     100
gacgccagaggacattattacattacagtagttcgccgctagatgacaaa     150
cgacatccttaccgatatgagatgtgcaaagctacataatggcaacaagc     200
gttatgaacagccttgtctttacgaccacagaaaagccgtattagagctc     250
ttcagctgcaaaattttcttctaatatgatgcaaagccatcaaaaatcat     300
gcatagttatgaaatacctgatgaaacgcttcgagttcgtgctcaagaaa     350
ttactgaaaggttaccgagaagaaaaatatctatgagacacgataaggcc     400
ccttctgaatccattgtcctgggcttgttcattctatttaccacttaaaa     450
ttgatcctttcaaaggaatttttttctatttccaatagtatatttgtaca     500
aaaactacaaaaatggataaaaaataacagtaatttgtgactactgtaaa     550
tatcactgatttggattttgtaatgagtactgctcatgcccatgccgatg     600
caagtggatcataaattttactaaacgatattcgataatgcgccaagcct     650
ttataaggaactcaaaataacccatatggacagtttcagaaggccaaata     700
acgatcaaggacattcactcatgtttttcaaaggcgaagagtgtaaaatt     750
ttcttctatatagttcgaatattttatcttataaatttcagtcgtcattt     800
```

FIG. 48

Utility3s

nt: SEQ ID NO: 44

```
tctgttaaagcagctgcgttgttcttgcactctgcacgtaaatgacgacg      50
gccaatgtaaaaatagcagtcactgcaggcctcttggattgtacccgtaa     100
ttacagcaattgccattagatttacgaagtcatttaattaaatggtttga     150
tttatctccttaaatgcttccagaaatgggactagacttttttcactcaa     200
acctgttcacaattattgctctttttcaattataaggtaaacaaggccat     250
ctatcagcaacacagtgctcgcatttttttaattaaactatataaaaccaa     300
ctatttgtggttgcgacttcacttttttgttgaattactacccaatcatta     350
atattgaagatgtgagatcatagatttattggctttgggcatctcaaatc     400
ccaagaggtcattttaaccaacaacattttaaaaagtagatttgtctgcc     450
tcagctatgagatgcgcatgtccctagcatctcatatctggttatattat     500
tttttccacttggtgaatgttgaaaaaaacaccactcgtccaatttatca     550
gtttgcaggtctaatgtccttccctgttatttaaactgtatattgtaagc     600
atgtcttatcgaaacaacttactcagttgtccgaaaacaaaactgcaaat     650
tctgtgtgtattcacgtactagaatcctgtcaaattggatcttgatttaa     700
gcttttatagcaacgaactttgcatactaagttttttttgttaaccggaa     750
ctgccaagaagcattcagtaaaatacatcttcatcatttactgataatac     800
tcattcagactcatatcatactatttcgaattcattatacatcctcaaaa     850
```

FIG. 49

Negative1s

nt: SEQ ID NO: 45

```
gctaggatctatatgcgaatatatcacatatgtaaattataagctcatcg      50
caaaaccaaaaaaaaaaaaaattttcaataatttttcactaatcttcaaaa     100
acaaatggggtaacccgtacaagagttattaaaacccaaaatgacaaaat     150
cgcgacaattcaatcctacttaattagcaataacatactagcggtagagc     200
tactatcacatgttgaaccttgaatgctcaattcattgtactcaatactg     250
ctatcaaaagaaaaaaaatgtattaattatattcttgtcaaaatcaattt     300
tacactataagaggaaaatgttcttcagtcctagtaacattagttttctc     350
cctttgctagagactttacataatatcctagaaggtaaaattcgataata     400
cagcagtaaagtcgtatattggtagcaatccttggtgacgctgacttttt     450
ttttttgtaattttattgtttagttcatgataaaaaacttcaaatcactt     500
ttaatctggtagacagagaaaacaaatcgaaacgaaaatagagaactacg     550
aataaaaaaatataagtggagaagatcgtcactacgcattaaacaatatt     600
gatcgctcaatgccagtactgcgcgtaaaagtttagtaacttaacgattt     650
aggcacaatttgagaaaaatttcgccctgcagtaagtatgttattcagta     700
cgatataaagctgaggttttatgct                              725
```

FIG. 50

Negative2s

nt: SEQ ID NO: 46

```
ggaagttgaagtttttcaaaaatttgctgtaaaatttgaccgatttgtga      50
gattcttcctggctgtcagaatatggggccgtagtatattgtcagacctg     100
ttcctttaagaggtgatggtgataggcgttgagtatgtgtagtgtttgac     150
ccgagggtatggttttcacaagtactgcgcactgtattgtgaaagcagct     200
tcggggtgcgtgattaaaaaatgcgaccaagaataaacaggtacatcata     250
acaagggccatttgaattgcatttatcaggatttgtaaccttgttctaaa     300
gaggcatcgtatagtttaagttcattttccacccaatttgatgacggtgt     350
ggaccttaacctattgtcttgaaatttaggttatctcttagatatcacat     400
gtgattaccccagtgaacgcgtataagcttacagaaaggaaaaccggttg     450
gctcagtcaaaactgttgcagatttgggctcccctgaatatttgagacat     500
ccctaaaatgaagagatatatacagctaattttgaatgaaaatttaaaat     550
tcgcaatgaacagtactagagatgagcttttgaagtcctttcaaattatt     600
tgttcttccagttgatatttttttatttatataccagtaccaaatataat     650
cttgccatacatttacctttttgaggttgttcaacggaaatccagtgtat     700
ttacacattcttggaaacccatcgcttataatacgaactaatttatttat     750
gaacaaaggctttggaaaagtatccctacttttttacgacgctaaatcatg     800
atacgaaactttaggaagattaacagtcactccataaaatcagaaagtat     850
tcgctaatagtggaaagaaatggttatataa                        881
```

FIG. 51